Psychological Methods
2012, Vol. 17, No. 1, 100–119

1082-989X/12/$12.00 DOI: 10.1037/a0025385

Clusterwise Simultaneous Component Analysis for Analyzing Structural Differences in Multivariate Multiblock Data

Kim De Roover and Eva Ceulemans
Katholieke Universiteit Leuven

Marieke E. Timmerman
University of Groningen

Kristof Vansteelandt, Jeroen Stouten, and Patrick Onghena
Katholieke Universiteit Leuven

Many studies yield multivariate multiblock data, that is, multiple data blocks that all involve the same set of variables (e.g., the scores of different groups of subjects on the same set of variables). The question then rises whether the same processes underlie the different data blocks. To explore the structure of such multivariate multiblock data, component analysis can be very useful. Specifically, 2 approaches are often applied: principal component analysis (PCA) on each data block separately and different variants of simultaneous component analysis (SCA) on all data blocks simultaneously. The PCA approach yields a different loading matrix for each data block and is thus not useful for discovering structural similarities. The SCA approach may fail to yield insight into structural differences, since the obtained loading matrix is identical for all data blocks. We introduce a new generic modeling strategy, called clusterwise SCA, that comprises the separate PCA approach and SCA as special cases. The key idea behind clusterwise SCA is that the data blocks form a few clusters, where data blocks that belong to the same cluster are modeled with SCA and thus have the same structure, and different clusters have different underlying structures. In this article, we use the SCA variant that imposes equal average cross-products constraints (ECP). An algorithm for fitting clusterwise SCA–ECP solutions is proposed and evaluated in a simulation study. Finally, the usefulness of clusterwise SCA is illustrated by empirical examples from eating disorder research and social psychology.

Keywords: multigroup data, multilevel data, principal component analysis, simultaneous component analysis, clustering

In the behavioral sciences, many studies yield multivariate multiblock data, that is, multiple data blocks that all involve the same set of variables. For an example, one can think of data from different groups of subjects that are measured on the same variables or of data from multiple subjects that have scored the same variables on multiple measurement occasions (the latter type of data are sometimes called multioccasion–multisubject data; see Kroonenberg, 2008). In the first case the different groups constitute the separate data blocks, in the second case the different subjects.

The question then arises whether the same structure underlies each data block. For example, in emotion psychology, there has been a long-lasting debate about the structure of emotions (e.g., Ekman, 1999; Fontaine, Scherer, Roesch, & Ellsworth, 2007; Russell & Barrett, 1999). In this debate, many cross-cultural psychologists argue that the structure may differ between cultures (Eid & Diener, 2001; Fontaine, Poortinga, Setiadi, & Markam, 2002; MacKinnon & Keating, 1989; Rodríguez & Church, 2003). In a similar vein, between subjects, structural differences in the time-varying experience of emotions can be expected. For example, studies have indicated that subjects differ with respect to so-called emotional granularity (Barrett, 1998; Tugade, Fredrickson, & Barrett, 2004). For subjects with a low emotional granularity, negative or positive emotional states are highly correlated across time. Subjects with high emotional granularity experience emotions in a more differentiated manner, distinguishing between a variety of positive and negative emotions, rather than feeling overall positive or negative.

To explore the similarities and differences in the structure of multivariate multiblock data, two approaches have been proposed within the component analysis literature. A first approach is to perform a separate principal component analysis (PCA; Jolliffe, 1986; Meredith & Millsap, 1985; Pearson, 1901) on each data block, which summarizes the information in the data block by reducing the variables to a few components. For instance, when analyzing cross-cultural emotion data, where each data block holds the emotion scores of inhabitants of a particular country, a PCA is

This article was published Online First October 3, 2011.

Kim De Roover, Eva Ceulemans, Jeroen Stouten, and Patrick Onghena, Faculty of Psychology and Educational Sciences, Katholieke Universiteit Leuven, Leuven, Belgium; Marieke E. Timmerman, Heymans Institute of Psychology, University of Groningen; Kristof Vansteelandt, Universitair Psychiatrisch Centrum, Katholieke Universiteit Leuven.

The research reported in this article was partially supported by the Fund for Scientific Research–Flanders (Belgium), Project No. G.0477.09, in an award to Eva Ceulemans, Marieke Timmerman, and Patrick Onghena, and by the Research Council of Katholieke Universiteit Leuven (GOA/10/02).

Correspondence concerning this article should be addressed to Kim De Roover, Department of Educational Sciences, Katholieke Universiteit Leuven, Andreas Vesaliusstraat 2, B-3000 Leuven, Belgium. E-mail: Kim.DeRoover@ped.kuleuven.be

conducted for each country separately (e.g., Fontaine et al., 2002). In this analysis, each data block $\mathbf{X}_i$ is decomposed into a component score matrix $\mathbf{F}_i$, containing the scores of the inhabitants on the components, and a loading matrix $\mathbf{B}_i$, indicating the extent to which the scores on the emotion variables are determined by the respective components. Because a separate loading matrix is obtained for each data block, this approach leaves plenty of freedom to trace differences in the underlying structure of the different data blocks. To gain further insight, van de Vijver and Leung (1997) proposed to rotate the loading matrices of the different data blocks toward each other and calculate Tucker congruence coefficients (Tucker, 1951) between them. However, when the loadings differ substantially, such coefficients do not reveal potential similarities across data blocks. Alternatively, one may compare the component loadings of the variables directly. Because of the large amount of information, this is often not very insightful either. Especially when the number of data blocks becomes large, it is practically infeasible to trace differences and similarities.

A second approach is simultaneous component analysis (SCA; Kiers, 1990; Kiers & ten Berge, 1994; Millsap & Meredith, 1988; Timmerman & Kiers, 2003; Van Deun, Smilde, van der Werf, Kiers, & Van Mechelen, 2009). In an SCA analysis, all data blocks are reduced simultaneously, based on the assumption that the same components underlie the different data blocks and thus that the same loading matrix can be used to reconstruct these data. By applying SCA to our example from cross-cultural emotion research, one common loading matrix would be obtained for all cultures. Since a common loading matrix is imposed, this approach is much more parsimonious than performing a PCA on each data block separately. However, this approach often leaves little room to find structural differences between data blocks.

One may conclude that both PCA and SCA may fail to yield insight, because in the separate PCA approach structural similarities are hard to trace, and in SCA structural differences may be difficult to detect. Therefore, we introduce a new generic modeling strategy called *clusterwise SCA*, which comprises the separate PCA approach and the SCA approach as special cases. Clusterwise SCA is based on the same principle as clusterwise linear regression, namely, that the data contain a few clusters, where each cluster has a different underlying structure (Brusco, Cradit, Steinley, & Fox, 2008; DeSarbo, Oliver, & Rangaswamy, 1989; Späth, 1979, 1982). More specifically, the key idea behind clusterwise SCA is that the different data blocks form a limited number of mutually exclusive clusters, where data blocks that belong to the same cluster can be reconstructed by means of the same loadings, whereas data blocks that belong to different clusters have a different underlying component structure and thus imply a different loading matrix. In this article, the data blocks within each cluster are modeled with the SCA variant that imposes equal average cross-products constraints (SCA–ECP; Timmerman & Kiers, 2003). Therefore, the proposed modeling strategy will be called *clusterwise SCA–ECP*.

The remainder of this article is organized as follows: In the following section, the separate PCA and SCA–ECP approaches are recapitulated, followed by the introduction of the new clusterwise SCA–ECP model and a comparison of the latter model to other approaches for detecting structural differences and similarities in multivariate multiblock data. Next, the Data Analysis section describes the aim of and an algorithm for clusterwise SCA–ECP analysis and proposes a model selection procedure. In the fourth section, the performance of the algorithm is evaluated in three simulation studies. Then the model is illustrated with an application to time series data from eating disorder patients and an application to data from social psychology, with subjects being nested in experimental conditions. Finally, directions for future research are discussed.

Model

Data Structure and Preprocessing

Clusterwise SCA–ECP can be applied to all kinds of multivariate multiblock data, where *multivariate* indicates that multiple variables are involved and *multiblock* implies that the data can be divided in separate data blocks according to the hierarchical structure of the data. Different configurations are possible. For instance, the data blocks may represent different groups (e.g., cultures) where the observations within the data blocks stem from different individuals that belong to these groups, or the data blocks may represent individuals with different time points constituting the observations within the blocks (i.e., multioccasion–multisubject data). More formally, clusterwise SCA–ECP requires I data blocks $\mathbf{X}_i$ ($N_i \times J$) that contain scores on J variables, where the number of observations N_i ($i = 1, \ldots, I$) in each data block may differ between data blocks, subject to the restriction that N_i is larger than the number of components to be fitted (but preferably larger than J, for the sake of stable model estimates). These I data blocks can be concatenated into a N (observations) $\times$ J (variables) data matrix $\mathbf{X}$, where $N = \sum_{i=1}^{I} N_i$. A graphical presentation of the data structure is given in Figure 1.

Note that three-way three-mode data (for an introduction, see Kroonenberg, 2008) are a special case of multivariate multiblock data, in which all the groups consist of the same subjects, or in which all the subjects are measured on the same time points. Moreover, note that a distinction can be made between multigroup and multilevel multiblock data, based on whether the data blocks are considered fixed or random respectively (Timmerman, Kiers, Smilde, Ceulemans, & Stouten, 2009). For instance, if each data block contains the scores of a single subject, the data blocks are fixed when one is interested only in the subjects in the study and random when one wants to generalize the conclusions toward a larger population of subjects. As clusterwise SCA–ECP is a deterministic method (i.e., no distributional assumptions are made), it is applicable to both multigroup and multilevel data.

Clusterwise SCA–ECP is designed for tracing similarities and differences in the within structure of the data (i.e., the correlational structure within each of the data blocks). Therefore, the differences between the data blocks in the means of the variables, which is often called the between structure, should be removed from the data. This is achieved by centering the variables per data block. Moreover, arbitrary differences between the variables in measurement scale are usually eliminated in component analysis by scaling the data. In the case of multivariate multiblock data, two scaling options have been advocated. The first option is autoscaling (Kiers & ten Berge, 1994), which implies that each variable is rescaled per data block, such that the sum of squares per variable is equal

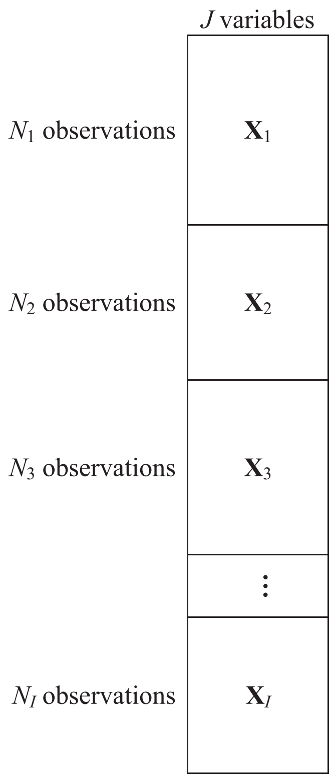

Figure 1. Graphical presentation of multivariate multiblock data.

to the number of observations N_i for the data block in question. This leads to a variance of 1 per variable for each data block. Note that when the centering per data block is combined with autoscaling, the preprocessing is equivalent to calculating z scores within each data block. The second option is to rescale across the data blocks (Timmerman & Kiers, 2003), implying a variance of 1 per variable across all data blocks, so that differences between the data blocks in variability are preserved. In this article we have chosen the first option, to focus on the correlational differences between the data blocks rather than the differences in variances. In what follows, we assume each data block $\mathbf{X}_i$ to be centered and autoscaled.

PCA on Each of the Data Blocks Separately

Applying PCA to each data block $\mathbf{X}_i$ ($i = 1, \ldots, I$) separately implies that the underlying structure is allowed to differ across data blocks in that the data blocks may be characterized by different loading matrices $\mathbf{B}_i$. Formally, this model can be written as follows:

$$\mathbf{X}_i = \mathbf{F}_i\mathbf{B}_i' + \mathbf{E}_i, \tag{1}$$

where $\mathbf{F}_i$ ($N_i \times Q$) denotes the component score matrix containing the Q component scores for observation $1, \ldots, N_i$, where the number of components Q is assumed to be the same across data blocks, $\mathbf{B}_i$ ($J \times Q$) denotes the loading matrix for the ith data block, and $\mathbf{E}_i$ ($N_i \times J$) denotes the matrix of residuals. Note that since the variables are standardized per data block, the loadings in $\mathbf{B}_i$ equal the correlations between the respective components and variables, in case of orthogonal components.

The loading matrices $\mathbf{B}_i$ ($i = 1, \ldots, I$) of the PCA solutions may be orthogonally or obliquely rotated, provided that such a transformation is compensated for in the component score matrices $\mathbf{F}_i$. Therefore, standard rotational procedures (e.g., varimax; Kaiser, 1958) can be applied to obtain solutions that are easier to interpret. When an oblique rotation is used, the components become correlated to some extent. Consequently, the loadings may not be interpreted as correlations but as weights that indicate the extent to which each variable is influenced by the respective components.

SCA–ECP

SCA differs from the separate PCA approach in that the underlying components are assumed to be the same across the data blocks, implying that the loadings of the variables are identical. Specifically, the SCA–ECP model (Kiers & ten Berge, 1994; Timmerman & Kiers, 2003) is given by

$$\mathbf{X}_i = \mathbf{F}_i\mathbf{B}' + \mathbf{E}_i, \tag{2}$$

where $\mathbf{F}_i$ ($N_i \times Q$) denotes the component score matrix of the ith data block; $\mathbf{B}$ ($J \times Q$) denotes the loading matrix, which is identical for all data blocks and therefore does not have an index i; and $\mathbf{E}_i$ ($N_i \times J$) denotes the matrix of residuals. In the SCA–ECP model, the component score matrices $\mathbf{F}_i$ are constrained in that the variances and the correlations of the component scores in $\mathbf{F}_i$ are restricted to be equal across data blocks (with all the component correlations being zero if the components are orthogonal), that is, $N_i^{-1}\mathbf{F}_i'\mathbf{F}_i = \mathbf{\Phi}$. To partly identify the solution, the variances of the component scores are fixed at 1, implying that the diagonal elements of $\mathbf{\Phi}$ are set to 1.

Note that Timmerman and Kiers (2003) also described three less restrictive variants of the SCA model, for which the variances and/or correlations of the component scores in $\mathbf{F}_i$ are allowed to vary across data blocks. Differences in the component variances and/or correlations may reveal some structural differences between the data blocks. For instance, when some of the data blocks have hardly any variance on a certain component, this may indicate that the component concerned does not underlie those data blocks. In that case, a rotation can be applied to distinguish such "distinctive" components from the common components more clearly (see DISCO–SCA; Schouteden, Van Deun, Van Mechelen, & Pattyn, 2010). However, in practice, differences in component variances (or correlations) are usually gradual rather than distinct, making it hard to deduce structural differences between the data blocks. On top of that, SCA–ECP imposes equality of the component variances and correlations across data blocks and thus rules out tracing structural differences. In the next section, we motivate why we selected the SCA–ECP variant for the clusterwise SCA.

As is the case for PCA solutions, the components of an SCA–ECP solution can be freely rotated without altering the fit of the solution. Specifically, the loading matrix $\mathbf{B}$ can be transformed by multiplying it by any rotation matrix, provided that such a transformation is compensated for in the component score matrices $\mathbf{F}_i$ ($i = 1, \ldots, I$).

Clusterwise SCA–ECP

As stated in the introduction, the key idea behind clusterwise SCA–ECP is that the different data blocks fall apart into K mutually exclusive clusters. The data blocks that belong to the same cluster are driven by the same processes, whereas data blocks that belong to different clusters come about through different mechanisms. Clusterwise SCA–ECP deals with these differences in underlying structure by partitioning the I data blocks into K clusters and modeling the data blocks within each cluster by an SCA–ECP model, assuming the number of components Q to be the same across the clusters.

In this article, the goal is to capture the structural differences between the data blocks with the clustering. To ensure that data blocks with a different within structure will be allocated to different clusters, the SCA–ECP model is imposed for each cluster. Any alternative SCA model would leave room for structural differences between the data blocks within a cluster, such as differences in variances of the component scores.

Formally, the model equation of clusterwise SCA–ECP is given by

$$\mathbf{X}_{\mathrm{i}} = \sum_{\mathrm{k}=1}^{\mathrm{K}} p_{\mathrm{ik}} \mathbf{F}_{\mathrm{i}}^{(\mathrm{k})} \mathbf{B}^{(\mathrm{k})\prime} + \mathbf{E}_{\mathrm{i}} = \sum_{\mathrm{k}=1}^{\mathrm{K}} p_{\mathrm{ik}} \mathbf{F}_{\mathrm{i}} \mathbf{B}^{(\mathrm{k})\prime} + \mathbf{E}_{\mathrm{i}}, \tag{3}$$

where K is the number of clusters; p_{ik} denotes the entries of the binary partition matrix $\mathbf{P}$ ($I \times K$), which equal 1 when data block i is assigned to cluster k and 0 otherwise; $\mathbf{F}_i^{(k)}$ ($N_i \times Q$) denotes the component score matrix of data block i when assigned to cluster k; $\mathbf{B}^{(k)}$($J \times Q$) denotes the loading matrix of cluster k ($k = 1, \ldots, K$); and $\mathbf{E}_i$ ($N_i \times J$) denotes the matrix of residuals. The loading matrices $\mathbf{B}^{(k)}$ are shared by all data blocks that belong to a particular cluster. As can be seen in the left part of Equation 3, the index k in $\mathbf{F}_{\mathrm{i}}^{(\mathrm{k})}$ is mostly omitted in the remainder of this article, because each data block $\mathbf{X}_i$ is assigned to one cluster only. Note that since the parameter estimates of an SCA–ECP solution have rotational freedom, the components of a clusterwise SCA–ECP solution can also be freely rotated within each cluster without altering the fit of the solution.

To illustrate the clusterwise SCA–ECP model, we make use of the hypothetical data matrix $\mathbf{X}$ in Table 1. $\mathbf{X}$ contains the scores of four subjects on six variables pertaining to emotions and physical activity at 8, 9, 7, and 10 measurement occasions, respectively. This data matrix can be perfectly reconstructed by the clusterwise SCA–ECP solution with two clusters and two components, of which the partition matrix, cluster loading matrices, and component score matrices are presented in Tables 2, 3, and 4, respectively.

From Table 2, it can be read that the four subjects fall apart into two clusters, where Subjects 1 and 4 belong to the first cluster and Subjects 2 and 3 to the second. To trace the structural differences between these two clusters, we inspect the cluster loading matrices $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ in Table 3. For both clusters, a positive affect component and a negative affect component are found. The difference between the two clusters lies in how positive and negative emotions are related to physical activity. Specifically, $\mathbf{B}^{(1)}$ shows that for the subjects in Cluster 1, there is a relation between physical activity and positive affect, with both being unrelated to negative affect, whereas $\mathbf{B}^{(2)}$ reveals that for the subjects of Cluster 2, being physically active is related to negative affect and not to positive affect. From the component score matrices $\mathbf{F}_i$ in Table 4, it can be derived how positive or negative the subjects feel on the different measurement occasions. Note that the scores in Table 1 can be reconstructed by combining the matrices in Tables 2, 3, and 4 according to Equation 3. For example, from Table 1 it can be read that Subject 1 has a score of −1.4 on the variable *happy*. By means of Equation 3, we can reconstruct this score as follows: $[-1.4 \quad -1.1] * [1 \quad 0]'$, where * denotes the inner product of the vectors, $[-1.4 \quad -1.1]$ are the component scores of the first observation of the subject in question, and $[1 \quad 0]$ are the loadings of *happy* for the first cluster, to which the subject belongs.

Finally, from Equation 3 it is clear that the clusterwise SCA–ECP model is a generic modeling strategy for tracing differences and similarities in underlying structure, in that it comprises the separate PCA and SCA–ECP approaches as special cases. Specifically, when the number of clusters K is set to 1, the model reduces to a regular SCA–ECP model. In case the number of clusters K equals the number of data blocks I, the model boils down to a separate PCA on each data block.

Relations to Existing Models

In the past decades, several models and associated algorithms have been developed for tracing differences in the underlying structure of multivariate data. In this section, we focus on methods for multiblock data, discarding a number of techniques that combine clustering and dimension reduction in the context of single block data, such as reduced K-means (Bock, 1987; de Soete & Carroll, 1994), factorial K-means (Timmerman, Ceulemans, Kiers, & Vichi, 2010; Vichi & Kiers, 2001), mixtures of factor analyzers (McLachlan & Peel, 2000), mixture structural equation modeling (SEM; Dolan & van der Maas, 1998; Jedidi, Jagpal, & DeSarbo, 1997; Yung, 1997), and high-dimensional data clustering (Bouveyron, Girard, & Schmid, 2007); and three-way three-mode data, such as Tucker3 clustering (Rocci & Vichi, 2005) and three-way factorial K-means (Vichi, Rocci, & Kiers, 2007). All existing techniques differ from clusterwise SCA–ECP in at least one of the following respects: Is the method exploratory or confirmatory? Is the model deterministic or stochastic (i.e., are assumptions made about the underlying probability distributions?)? What kind of data are required: two-mode data or one-mode data? What type of data reduction is performed?

Within the family of deterministic models, clusterwise SCA–ECP is related to points-of-view analysis (Tucker & Messick, 1963). This method was proposed to handle multiple one-mode (dis)similarity data blocks. The data blocks are clustered, and for each cluster multidimensional scaling is performed (Kruskal & Wish, 1978). It can be concluded that clusterwise SCA–ECP differs from this method with respect to the kind of data that are dealt with and the type of data reduction, as clusterwise SCA–ECP deals with two-mode instead of one-mode data blocks and uses SCA–ECP instead of multidimensional scaling for the data reduction.

Within the stochastic framework, clusterwise SCA–ECP is mainly related to multigroup SEM (Jöreskog, 1971; Kline, 2004; Sörbom, 1974) and multigroup exploratory factor analysis (EFA; Dolan, Oort, Stoel, & Wicherts, 2009; Hessen, Dolan, & Wicherts, 2006) on the one hand and multilevel latent class analysis (Ver-

Table 1
*Hypothetical Data Matrix **X** With the (Rounded Off) Scores of Four Subjects on Six Variables Measuring Emotions and Physical Activity*

Observation	Happy	Pleased	Sad	Ashamed	Moving	Sporting
			Subject 1			
1	−1.4	−1.4	−1.1	−1.1	−1.4	−1.4
2	0.5	0.5	2.4	2.4	0.5	0.5
3	−0.2	−0.2	−0.6	−0.6	−0.2	−0.2
4	−0.2	−0.2	0.7	0.7	−0.2	−0.2
5	1.4	1.4	−0.2	−0.2	1.4	1.4
6	0.3	0.3	0.9	0.9	0.3	0.3
7	0.2	0.2	−0.8	−0.8	0.2	0.2
8	1.6	1.6	−1.4	−1.4	1.6	1.6
			Subject 2			
1	2.6	2.6	−0.8	−0.8	−0.8	−0.8
2	−0.7	−0.7	0.7	0.7	0.7	0.7
3	0.2	0.2	0.8	0.8	0.8	0.8
4	−0.1	−0.1	−0.2	−0.2	−0.2	−0.2
5	−1.9	−1.9	0.2	0.2	0.2	0.2
6	−0.4	−0.4	−1.2	−1.2	−1.2	−1.2
7	−1.8	−1.8	−1.1	−1.1	−1.1	−1.1
8	0.8	0.8	0.1	0.1	0.1	0.1
9	−0.9	−0.9	0.7	0.7	0.7	0.7
			Subject 3			
1	−0.2	−0.2	0.1	0.1	0.1	0.1
2	−2.1	−2.1	−0.5	−0.5	−0.5	−0.5
3	−0.8	−0.8	0.3	0.3	0.3	0.3
4	1.4	1.4	−0.6	−0.6	−0.6	−0.6
5	−1.1	−1.1	0.5	0.5	0.5	0.5
6	1.0	1.0	0.7	0.7	0.7	0.7
7	0.1	0.1	1.7	1.7	1.7	1.7
			Subject 4			
1	1.1	1.1	1.4	1.4	1.1	1.1
2	−0.3	−0.3	−2.0	−2.0	−0.3	−0.3
3	0.7	0.7	−0.2	−0.2	0.7	0.7
4	−2.1	−2.1	−1.2	−1.2	−2.1	−2.1
5	−0.4	−0.4	2.9	2.9	−0.4	−0.4
6	−0.8	−0.8	0.8	0.8	−0.8	−0.8
7	−1.6	−1.6	1.4	1.4	−1.6	−1.6
8	0.5	0.5	−1.1	−1.1	0.5	0.5
9	0.3	0.3	−0.5	−0.5	0.3	0.3
10	0.0	0.0	−0.3	−0.3	0.0	0.0

munt, 2003, 2008a) and multilevel mixture factor analysis (Varriale & Vermunt, 2010; Vermunt, 2008b) on the other hand.

Multigroup SEM is a confirmatory method for multiblock data that imposes a structural equation model (Haavelmo, 1943; Kline, 2004) for each data block. Similarly, multigroup EFA is an exploratory method, in which an exploratory common factor model (Lawley & Maxwell, 1962) is estimated for each data block. The factor loading structure can be constrained to be the same for each data block (as in SCA) or different for each data block (as in separate PCA). To determine whether the factor loading structure is invariant over the blocks, the fit of the constrained and unconstrained model are compared. Thus, multigroup SEM and multigroup EFA differ from clusterwise SCA–ECP in that these methods do not perform a clustering of the data blocks. Moreover, in multigroup SEM and multigroup EFA, it is assumed that the observations within a particular data block are independent. Therefore, modeling multioccasion–multisubject data with multigroup SEM or EFA would not be correct, due to the serial dependencies between successive measurements. Thus, clusterwise SCA–ECP has as a major advantage over multigroup SEM and EFA that no assumptions about within-block dependencies are made. When these dependencies exist, they may be captured in the component scores.

Multilevel latent class analysis as well as multilevel mixture factor analysis implies a clustering of the data blocks. However, a crucial difference between the latter techniques and clusterwise SCA–ECP is that in multilevel latent class analysis and multilevel mixture factor analysis the clustering is based on between-block differences in variable means, thus capturing the between structure, whereas in clusterwise SCA–ECP the clustering of the data blocks is performed according to between-block differences in within-block structure.

From this overview of related models, it can be concluded that no other model is available for two-mode multivariate multiblock data that clusters the data blocks based on the within-block structure.

Table 2
*Partition Matrix **P** of the Clusterwise SCA–ECP Decomposition With Two Clusters and Two Components of **X*** in Table 1

Subject	Cluster 1	Cluster 2
1	1	0
2	0	1
3	0	1
4	1	0

Note. The 1s indicate to which cluster each subject is assigned. SCA–ECP = simultaneous component analysis with equal average cross-products constraints.

Data Analysis

Aim of Clusterwise SCA–ECP Analysis

For a given number of clusters K and components Q, the aim of a clusterwise SCA–ECP analysis is to find the partition matrix $\mathbf{P}$, the component score matrices $\mathbf{F}_i$, and the loading matrices $\mathbf{B}^{(k)}$ that minimize the loss function:

$$L = \sum_{i=1}^{I} \sum_{k=1}^{K} p_{ik} \, || \, \mathbf{X}_i - \mathbf{F}_i \mathbf{B}^{(k)\prime} \, ||^2. \tag{4}$$

Note that on the basis of the loss function value L, the percentage of variance accounted for (VAF) can be computed as follows:

$$\mathrm{VAF}(\%) = \frac{||\mathbf{X}||^2 - L}{||\mathbf{X}||^2} \times 100. \tag{5}$$

Algorithm

Ideally, one would wish to develop a clusterwise SCA–ECP algorithm that returns a globally optimal solution. To this end, the algorithm should sieve through all possible partitions of the data blocks in a smart way. However, for the related K-means clustering problem (MacQueen, 1967), this becomes very time-consuming if the data set contains a large number of observations (Aloise, Hansen, & Liberti, 2010; Brusco, 2006), because the number of possible partitions grows very large. Our problem requires additional computations, since data reduction is performed within each cluster, thus a search of all possible partitions seems not feasible for clusterwise SCA–ECP. Therefore, we developed a fast relocation algorithm that can handle large numbers of data blocks but may end in a local minimum.

More specifically, to obtain a (K,Q) clusterwise SCA–ECP solution that minimizes the loss function (Equation 4), an alternating least squares procedure is used, which consists of four steps:

1. Randomly initialize partition matrix $\mathbf{P}$: Initialize the partition matrix $\mathbf{P}$ by randomly assigning the I data blocks to one of the K clusters, where each data block has an equal probability of being assigned to each cluster. If one of the clusters is empty, repeat this procedure until all clusters contain at least one element.

2. Estimate the SCA–ECP model for each cluster: Estimate the $\mathbf{F}_i$ and $\mathbf{B}^{(k)}$ matrices for each cluster $k = 1, \ldots, K$ by performing a rationally started SCA–ECP analysis (Timmerman & Kiers, 2003) on the $\mathbf{X}_i$ data blocks assigned to the kth cluster. To this end, the loading matrix $\mathbf{B}^{(k)}$ is rationally initialized, based on the singular value decomposition of the vertical concatenation of the data blocks within cluster k, denoted by $\mathbf{X}^{(k)}$. Next, an alternating least squares procedure is performed, in which $\mathbf{F}^{(k)}$ and $\mathbf{B}^{(k)}$ are iteratively reestimated, where $\mathbf{F}^{(k)}$ is the vertical concatenation of the component scores of all data blocks within cluster k. Specifically, $\mathbf{F}^{(k)}$ is (re)estimated by performing a singular value decomposition for all data blocks $\mathbf{X}_i$ that belong to cluster k: $\mathbf{X}_i\mathbf{B}^{(k)}$ is decomposed into $\mathbf{U}_i$, $\mathbf{S}_i$, and $\mathbf{V}_i$ with $\mathbf{X}_i\mathbf{B}^{(k)} = \mathbf{U}_i\mathbf{S}_i\mathbf{V}_i'$; a least squares estimate of $\mathbf{F}_i^{(k)}$ is given by $\mathbf{F}_i^{(k)} = \sqrt{N_i}\mathbf{U}_i\mathbf{V}_i'$ (ten Berge, 1993). $\mathbf{B}^{(k)}$ is updated by $\mathbf{B}^{(k)} = ((\mathbf{F}^{(k)\prime}\mathbf{F}^{(k)})^{-1}\mathbf{F}^{(k)\prime}\mathbf{X}^{(k)})'$. This way, cluster loading matrices $\mathbf{B}^{(k)}$ are obtained that "average" the underlying structure of the data blocks within a cluster. Note that due to the SCA–ECP restrictions on the component variances and correlations of each $\mathbf{F}_i^{(k)}$, the $\mathbf{F}^{(k)}$ and $\mathbf{B}^{(k)}$ matrices for each cluster k cannot be obtained directly by means of a singular value decomposition of $\mathbf{X}^{(k)}$.

3. Reestimate the partition matrix $\mathbf{P}$: For each data block, a component score matrix $\mathbf{F}_i^{(k)}$ is computed for each cluster k, by means of the same computations that are

Table 3
*Component Loading Matrices $\mathbf{B}^{(1)}$ and $\mathbf{B}^{(2)}$ of the Clusterwise SCA–ECP Decomposition With Two Clusters and Two Components of **X** in Table 1*

	Cluster 1		Cluster 2	
Variable	Positive affect and physical activity	Negative affect	Positive affect	Negative affect and physical activity
Happy	1	0	1	0
Pleased	1	0	1	0
Sad	0	1	0	1
Ashamed	0	1	0	1
Moving	1	0	0	1
Sporting	1	0	0	1

Note. SCA–ECP = simultaneous component analysis with equal average cross-products constraints.

Table 4
Component Score Matrices $\boldsymbol{F}_{\mathrm{i}}$ of the Clusterwise SCA–ECP Decomposition With Two Clusters and Two Components of $\boldsymbol{X}$ *in Table 1*

	Cluster 1		Cluster 2	
Observation	Positive affect and physical activity	Negative affect	Positive affect	Negative affect and physical activity
		Subject 1		
1	−1.4	−1.1		
2	0.5	2.4		
3	−0.2	−0.6		
4	−0.2	0.7		
5	1.4	−0.2		
6	0.3	0.9		
7	0.2	−0.8		
8	1.6	−1.4		
		Subject 2		
1			2.6	−0.8
2			−0.7	0.7
3			0.2	0.8
4			−0.1	−0.2
5			−1.9	0.2
6			−0.4	−1.2
7			−1.8	−1.1
8			0.8	0.1
9			−0.9	0.7
		Subject 3		
1			−0.2	0.1
2			−2.1	−0.5
3			−0.8	0.3
4			1.4	−0.6
5			−1.1	0.5
6			1.0	0.7
7			0.1	1.7
		Subject 4		
1	1.1	1.4		
2	−0.3	−2.0		
3	0.7	−0.2		
4	−2.1	−1.2		
5	−0.4	2.9		
6	−0.8	0.8		
7	−1.6	1.4		
8	0.5	−1.1		
9	0.3	−0.5		
10	0.0	−0.3		

Note. SCA–ECP = simultaneous component analysis with equal average cross-products constraints.

used for updating the component score matrices within SCA–ECP (see Step 2). The extent to which the data block fits in the different clusters is quantified by computing the following block-and-cluster-specific partition criterion: $L_{\mathrm{ik}} = \| \mathbf{X}_{\mathrm{i}} - \mathbf{F}_{\mathrm{i}}^{(k)}\mathbf{B}^{(k)\prime}\|^2$. Each data block is assigned to the cluster k for which L_{ik} is minimal. When one of the K clusters is empty after this procedure, the data block with the worst fit in its current cluster is moved to the empty cluster.

4. Steps 2 and 3 are repeated until convergence is reached, that is, until the decrease of the loss function value L (Equation 4) for the current iteration is smaller than the convergence criterion of 0.000001.

To reduce the probability of ending up in a local minimum, it is advised to use a multistart procedure with different random initializations of the partition matrix **P**. In the algorithm, a rational start is used for the SCA–ECP analysis within each cluster (see Step 2 of the algorithm). The results of a pilot study reveal that the use of multiple random starts rather than a single rational start does not improve the performance of the clusterwise SCA–ECP algorithm, and therefore we deem the use of multiple random starts to be unnecessary for the within-cluster SCA–ECP analyses. The described algorithm has been implemented in MATLAB R2010a and can be obtained freely from the first author. Moreover, user-friendly software for applying clusterwise SCA–ECP has already been developed (De Roover, Ceulemans, & Timmerman, 2011) and is available online for potential users (http://ppw.kuleuven.be/okp/software/MBCA/). It is also possible to run the software as a stand-alone application (i.e., without MATLAB).

Model Selection

When applying clusterwise SCA–ECP analysis, the underlying number of clusters K and components Q is usually unknown. To deal with this, clusterwise SCA–ECP solutions are estimated with several values for K and Q. Subsequently, a solution is selected on the basis of formal model selection techniques, interpretability, and stability of the solutions. As a formal model selection technique, one may consider using a generalization of the well-known scree test (Cattell, 1966), which aims at selecting a model with an optimal balance between fit and parsimony. Specifically, one may plot the percentage of VAF (Equation 5) of the different solutions against the number of components for each value of K. The number of clusters K is established by examining the general increase in fit (i.e., over the different numbers of components) that is obtained by adding a cluster and choosing the number of clusters, after which this general increase in fit levels off. Finally, considering the solutions with K clusters only, the number of components Q is determined for which it holds that adding more components does not significantly increase the fit to the data. The use of this model selection procedure is further illustrated below when applying it to empirical data sets. To evaluate the stability of the retained solution, one can examine how often (specific aspects of) the obtained partition shows up across the different random starts in the multistart procedure.

Simulation Studies

In this section, we first present a large simulation study in which the performance of the clusterwise SCA–ECP algorithm is evaluated when the correct number of clusters and components is known. Next, two smaller simulation experiments are discussed in which we investigate how the performance is affected by less favorable analysis conditions. Specifically, we examine the effect of using an incorrect number of clusters or components and the effect of the presence of masking variables in the data (i.e., variables in which the cluster structure is not reflected; see, e.g., Brusco & Cradit, 2001; Vichi & Kiers, 2001).

Simulation Study 1

Problem. The first simulation study is an extensive study in which the clusterwise SCA–ECP algorithm is evaluated with respect to goodness of fit, sensitivity to local minima, and goodness of recovery, under optimal conditions, that is, when the data to be analyzed are generated from a clusterwise SCA–ECP model with a known number of clusters K and components Q. Furthermore, we examine whether the performance of the algorithm is influenced by seven factors:

1. the number of data blocks,
2. the number of observations per data block,
3. the number of underlying clusters,
4. the number of underlying components,
5. the cluster size,
6. the amount of error on the data, and
7. the structure of the loading matrices of the different clusters.

These factors were chosen because their influence is often investigated in simulation studies on clustering algorithms and/or because they are theoretically interesting. Factors 1 and 2 pertain to sample size. On the basis of previous research (Brusco & Cradit, 2005; Hands & Everitt, 1987), we expect that the clusterwise SCA–ECP algorithm will perform better when more information is available (i.e., more data blocks and/or more observations per data block). With respect to Factors 3 and 4, which refer to the complexity of the underlying model, we hypothesize that the goodness of fit and recovery will decrease with increasing complexity (Brusco & Cradit, 2005; Milligan, Soon, & Sokol, 1983; Timmerman et al., 2010). Regarding Factor 5, cluster size, we conjecture that better results will be obtained when the clusters are of equal size (Brusco & Cradit, 2001; Milligan et al., 1983; Steinley, 2003). For Factor 6, amount of error, we expect that performance will deteriorate when the data contain more error (Brusco & Cradit, 2005). Finally, Factor 7, structure of the cluster loading matrices, was manipulated to study whether the performance deteriorates when the structures underlying the different clusters overlap more, where overlap is defined in terms of congruence between the loadings of the different clusters. Moreover, we expect ordinal interactions between (some of) these factors. Specifically, we conjecture that the worst results will be obtained with small data sets that contain, apart from a lot of error, many clusters of different sizes, of which the cluster loading matrices are highly congruent.

Design and procedure. In this simulation study, the number of variables J was fixed at 12. Furthermore, the seven factors that were introduced above were systematically varied in a complete factorial design:

1. the number of data blocks I at two levels: 20, 40;
2. the number of observations per data block N_i at three levels: $N_i \sim U[15;20]$, $N_i \sim U[30;70]$, $N_i \sim U[80;120]$, with U indicating a discrete uniform distribution between the given numbers;
3. the number of clusters K at three levels: 2, 3, 4;
4. the number of components Q at three levels: 2, 3, 4;
5. the cluster size, at three levels (see Milligan et al., 1983): equal (equal number of data blocks in each cluster), unequal with minority (10% of the data blocks in one cluster and the remaining data blocks distributed equally over the other clusters), and unequal with majority (60% of the data blocks in one cluster and the remaining data blocks distributed equally over the other clusters);

6. the error level e, which is the expected proportion of error variance in the data blocks $\mathbf{X}_i$: .00, .20, .40;

7. the structure of the loading matrices $\mathbf{B}^{(k)}$ of the different clusters at three levels: simple structure, random with low congruence, and random with high congruence.

With respect to the latter factor, at the simple structure level, the cluster loading matrices all had simple structure, with the set of variables constituting the different components varying across the different clusters. For instance, the cluster loading matrices for K and Q equal to four were constructed as follows:

$$\mathbf{B}^{(1)} = \begin{bmatrix} 1&0&0&0\\1&0&0&0\\1&0&0&0\\0&1&0&0\\0&1&0&0\\0&1&0&0\\0&0&1&0\\0&0&1&0\\0&0&1&0\\0&0&0&1\\0&0&0&1\\0&0&0&1 \end{bmatrix} \quad \mathbf{B}^{(2)} = \begin{bmatrix} 0&1&0&0\\1&0&0&0\\1&0&0&0\\0&0&1&0\\0&1&0&0\\0&1&0&0\\0&0&0&1\\0&0&1&0\\0&0&1&0\\1&0&0&0\\0&0&0&1\\0&0&0&1 \end{bmatrix}$$

$$\mathbf{B}^{(3)} = \begin{bmatrix} 1&0&0&0\\0&1&0&0\\1&0&0&0\\0&1&0&0\\0&0&1&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\\0&0&1&0\\0&0&0&1\\1&0&0&0\\0&0&0&1 \end{bmatrix} \quad \mathbf{B}^{(4)} = \begin{bmatrix} 1&0&0&0\\1&0&0&0\\0&1&0&0\\0&1&0&0\\0&1&0&0\\0&0&1&0\\0&0&1&0\\0&0&1&0\\0&0&0&1\\0&0&0&1\\0&0&0&1\\1&0&0&0 \end{bmatrix}$$

Random cluster loading matrices with low congruence were obtained by sampling the loadings uniformly between -1 and 1. Finally, random cluster loading matrices with high congruence were constructed as follows: First, a common base matrix was uniformly sampled between -1 and 1. Subsequently, the rows of this matrix were rescaled to have a sum of squares equal to .7. Next, for each cluster a new random matrix, uniformly sampled between -1 and 1 and rescaled to have a rowwise sum of squares equal to .3, was added to the base matrix to form the cluster-loading matrix.

To evaluate how much the resulting cluster loading matrices differ in each level of Factor 7, we set them orthogonally Procrustes rotated to one another (i.e., for each pair of cluster loading matrices, one was chosen to be the target matrix and the other was rotated toward the target matrix) and computed a congruence coefficient φ[1] (Tucker, 1951) for each pair of corresponding components in all pairs of $\mathbf{B}^{(k)}$ matrices. Subsequently, a grand mean of the obtained φ values was calculated, over the components and cluster pairs. The resulting φ values led to the conclusion that the differences between the cluster loading matrices are large at the random, low-congruence level (average φ across data sets = .41, SD = .09); rather small at the simple structure level (average φ = .71, SD = .07); and very small at the random, high-congruence level (average φ = .93, SD = .02). Therefore, the levels of this factor can also be labeled "low congruence," "medium congruence," and "high congruence," respectively.

For each cell of the factorial design, 50 data matrices $\mathbf{X}$ were generated, consisting of I data blocks $\mathbf{X}_i$. These data blocks were constructed as follows:

$$\mathbf{X}_i = \mathbf{F}_i\mathbf{B}^{(k)\prime} + \mathbf{E}_i, \tag{6}$$

where the component score matrix $\mathbf{F}_i$ and error matrix $\mathbf{E}_i$ were generated by randomly sampling entries from a standard normal distribution. The partition matrix $\mathbf{P}$ was generated by first computing the size of the different clusters and then randomly assigning the correct number of data blocks to the clusters. The cluster-loading matrices $\mathbf{B}^{(k)}$ were generated as described above (Factor 7). Subsequently, the error matrices $\mathbf{E}_i$ and the cluster loading matrices $\mathbf{B}^{(k)}$ were rescaled to obtain data that contain a proportion e of error variance (Factor 6). Finally, the resulting $\mathbf{X}_i$ matrices were standardized columnwise and were vertically concatenated into the matrix $\mathbf{X}$.

In total, 2 (number of data blocks) × 3 (number of observations per data block) × 3 (number of clusters) × 3 (number of components) × 3 (cluster size) × 3 (error level) × 3 (structure of cluster loading matrices) × 50 (replicates) = 72,900 simulated data matrices were generated. Each data matrix $\mathbf{X}$ was analyzed with the clusterwise SCA–ECP algorithm, with the correct number of clusters K and components Q. The algorithm was run 25 times, each time using a different random start, and the best solution was retained.

Results. We start by discussing the results concerning the goodness of fit, then we evaluate the sensitivity to local minima, and finally we examine the goodness of recovery.

Goodness of fit. On average the VAF (Equation 5) equals 99% (SD = 1.18), 84% (SD = 1.54), and 69% (SD = 2.74) when 0%, 20%, and 40% error variance is present in the data (Factor 6), respectively. Note that in the 0% error condition, the means and standard deviations of the simulated component score matrices will differ from 0 and 1, due to sampling fluctuations, implying that the simulated data do not perfectly comply with the SCA–ECP assumptions. This explains why these data sets cannot be fitted perfectly. The results for the 20% and 40% error conditions indicate that part of the error variance is fitted and thus that overfitting occurs. To gain more insight into when overfitting occurs, the amount of overfit was computed as the difference between the VAF by the estimated model and the VAF by the true model:

$$\text{overfit}(\%) = \text{VAF}_\text{M} - \text{VAF}_\text{T}$$

$$= \frac{\left(1 - \sum_{n=1}^{N}\sum_{j=1}^{J} \|\mathbf{X} - \mathbf{M}\|^2\right) - \left(1 - \sum_{n=1}^{N}\sum_{j=1}^{J} \|\mathbf{X} - \mathbf{T}\|^2\right)}{N \times J} \times 100, \tag{7}$$

[1] The congruence coefficient (Tucker, 1951) between two column vectors $\mathbf{x}$ and $\mathbf{y}$ is defined as their normalized inner product:

$$\varphi_{\mathbf{xy}} = \frac{\mathbf{x'y}}{\sqrt{\mathbf{x'x}}\sqrt{\mathbf{y'y}}}.$$

where **T** is the true data matrix (i.e., the data before error was added) and **M** contains the reconstructed scores of the best solution out of the 25 random runs. Note that **X** always differs from **T**, even in those cases where no error was added, because **X**, unlike **T**, is a preprocessed matrix. For all 72,900 data sets, the overfit value was larger than 0, implying that overfitting always occurs. Subsequently, an analysis of variance was performed with overfit as the dependent variable and the seven factors as independent variables. Finally, to examine which main and interaction effects have a large effect size, omega-squared proportions of total variance $\hat{\omega}^2$ (Hays, 1963; Olejnik & Algina, 2000) were computed. A large omega-squared was found for the main effect of the amount of error variance ($\hat{\omega}^2 = .68$), which implies that the error variance accounts for 68% of the variance in the overfit value. As expected, the overfit increases when more error variance is present in the data (see Figure 2A). The amount of error further interacts with the number of components ($\hat{\omega}^2 = .09$): the more components, the more overfitting (Figure 2A). Finally, a main effect of the number of observations per data block ($\hat{\omega}^2 = .14$) was revealed, implying that the amount of overfit is larger when less observations are available (see Figure 2B). Other effects are not discussed, because they account for less than 7.5% of the variance of the dependent variable.

***Sensitivity to local minima*.** The sensitivity of the clusterwise SCA–ECP algorithm to local minima should be evaluated by comparing the loss function value of the retained solution with that of the global minimum. However, the global minimum of the clusterwise SCA–ECP analysis of a data set **X** is unknown, because the simulated data do not perfectly comply with the clusterwise SCA–ECP assumptions, as was explained before.

Alternatively, we first evaluated whether the best fitting solution out of the 25 solutions from the multistart procedure yielded a higher loss function than the solution resulting from seeding the algorithm with the true $\mathbf{F}_i$, $\mathbf{B}^{(k)}$, and **P**. If this were the case, the multistart solution is a local minimum for sure. Such a local minimum was found for 1,230 out of the 72,900 simulated data matrices (1.69%). Out of these 1,230 data sets, 1,206 belong to the condition in which the cluster loading matrices are random with high congruence. Moreover, the majority (1,199) of these 1,206 data sets contain a lot of error variance (i.e., $e = .40$) and/or have a low number of observations per data block (i.e., between 15 and 20).

To investigate further the issue of local minima, we examined the proportion of random runs that had a loss function value that was equal to that of a proxy of the global minimum. This proportion will be called *global minimum proportion* in the remainder of the article. Note that the proxy is the best fitting solution resulting from seeding the algorithm with the true $\mathbf{F}_i$, $\mathbf{B}^{(k)}$, and **P** on the one hand and from the 25 random runs on the other hand. On average, the global minimum proportion equals .81 ($SD = .26$). An analysis of variance was performed with the global minimum proportion as the dependent variable, which revealed main effects of the number of observations per data block, of the number of clusters K, and of the type of cluster loading matrices, and an interaction effect of the number of observations per data block and the type of cluster-loading matrices. The effect of the number of observations per data block ($\hat{\omega}^2 = .11$) implies that having fewer observations per data block leads to a lower global minimum proportion (see Figure 3B). The effect of the number of clusters ($\hat{\omega}^2 = .17$) implies that the more clusters are present in the data, the lower the global minimum proportion (see Figure 3A). With respect to the effect of the amount of congruence between the cluster loading matrices ($\hat{\omega}^2 = .19$), the global minimum proportion is lower for the random-loading matrices with a high congruence (Figure 3B). The interaction of the latter effect with the number of observations per data blocks ($\hat{\omega}^2 = .09$) implies that the effect of the congruence between the cluster loading matrices is stronger when the number of observations per data block is lower, and vice versa (Figure 3B).

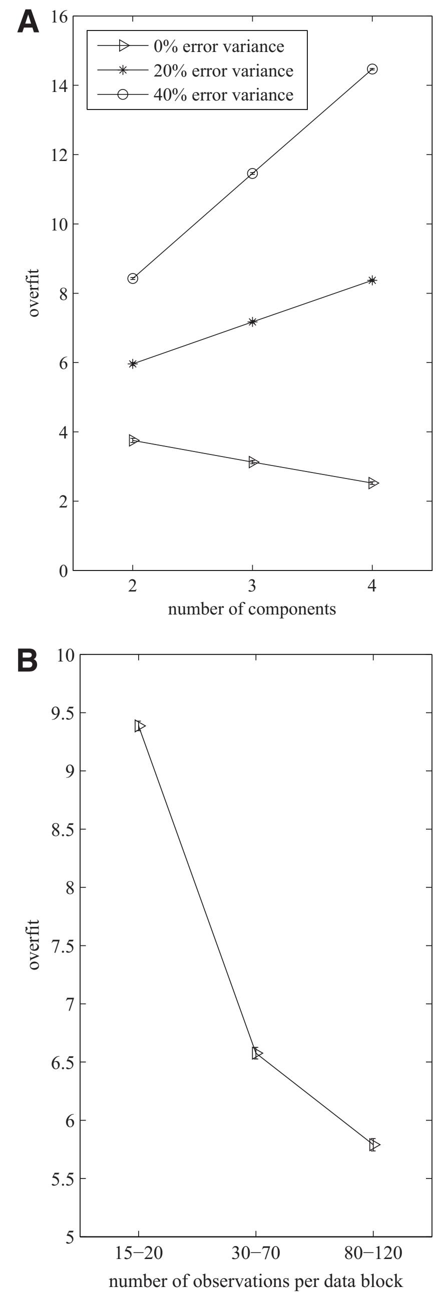

Figure 2. Mean overfit values and associated 95% confidence intervals as a function of the number of components and the amount of error variance ($e \times 100\%$; A) and as a function of the number of observations per data block (B).

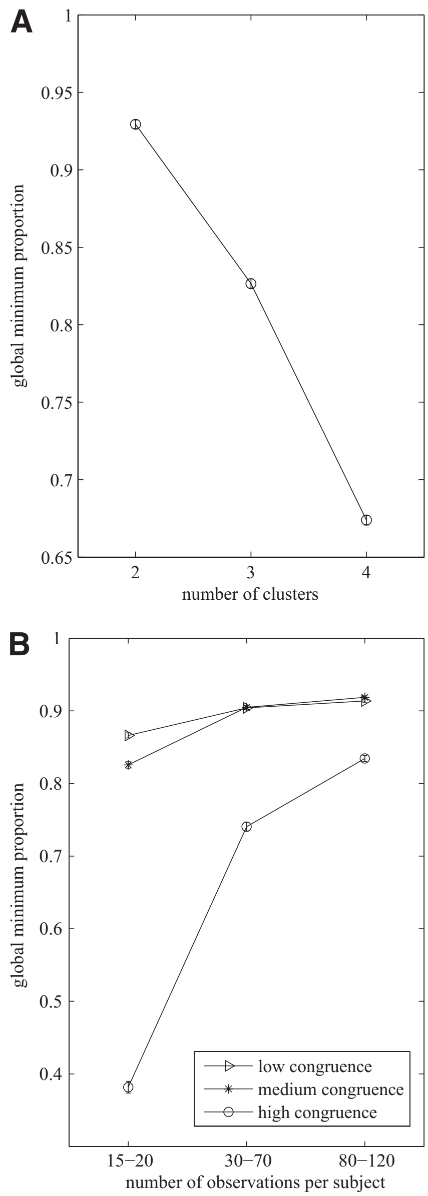

Figure 3. Mean values and associated 95% confidence intervals of the proportion of random runs with a loss function value equal to that of the proxy of the global minimum (i.e., global minimum proportion) as a function of the number of clusters (A) and as a function of the number of observations per data block and the congruence of the cluster loading matrices (B).

Goodness of recovery. The goodness of recovery will be evaluated with respect to (a) the clustering of the data blocks and (b) the cluster loading matrices.

Recovery of the clustering of the data blocks. To examine the goodness of recovery of the cluster membership of the data blocks, we computed the adjusted Rand index (ARI; Hubert & Arabie, 1985) between the true partition of the data blocks and the estimated partition. The ARI equals 1 if the two partitions are identical and 0 when the overlap between the two partitions is at chance level.

With an overall mean ARI of .98 (SD = .11) the clusterwise SCA–ECP algorithm appears to recover the clustering of the data blocks very well. An analysis of variance was performed with ARI as the dependent variable and the seven factors as independent variables. An interaction effect between the congruence of the cluster loading matrices and the number of observations per data block ($\hat{\omega}^2 = .12$) was found, which implies that the ARI is lower when the congruence between the cluster loading matrices is high, given that the number of observations per data block is low (i.e., between 15 and 20; see Figure 4).

Recovery of the cluster loading matrices. To evaluate the recovery of the cluster loading matrices, we obtained a goodness-of-cluster-loading-recovery (GOCL) statistic by computing congruence coefficients φ (Tucker, 1951) between the components of the true and estimated loading matrices and averaging across components and clusters as follows:

$$\mathrm{GOCL} = \frac{\sum_{k=1}^{K}\sum_{q=1}^{Q}\varphi(\mathbf{B}_{\mathrm{q}}^{(\mathrm{k})\mathrm{T}}, \mathbf{B}_{\mathrm{q}}^{(\mathrm{k})\mathrm{M}})}{KQ}, \tag{8}$$

with φ being the Tucker phi coefficient and $\mathbf{B}_{\mathrm{q}}^{(\mathrm{k})\mathrm{T}}$ and $\mathbf{B}_{\mathrm{q}}^{(\mathrm{k})\mathrm{M}}$ indicating the qth component of the true and estimated cluster loading matrices, respectively. The rotational freedom of the clusterwise SCA–ECP model was dealt with by an orthogonal Procrustes rotation of the estimated loading matrices toward the true loading matrices. To take the permutational freedom of the clusters into account, the permutation was chosen that maximizes the GOCL value. The GOCL statistic takes values between 0 (no recovery at all) and 1 (perfect recovery).

In our simulation study, the GOCL value varies between .86 and 1.00, with a mean value of .9969 (SD = .01), indicating an excellent recovery of the $\mathbf{B}^{(\mathrm{k})}$ matrices. An analysis of variance was performed with GOCL as the dependent variable, which revealed a main effect of the number of observations per data block ($\hat{\omega}^2 = .20$)—the lower the number of observations per data block, the worse the cluster loading matrices are recovered (see Figure 5)—and a main effect of the error level ($\hat{\omega}^2 = .14$)—the cluster loading matrices are more difficult to recover when the data contain more error (Figure 5). The main effects of error and the number of observations per data block are qualified by the interaction of these factors ($\hat{\omega}^2 = .08$): The effect of the number of observations per data block is stronger when the data contain more error (Figure 5).

Simulation Study 2

To investigate how the goodness of recovery is affected by performing clusterwise SCA–ECP with an incorrect number of clusters or components, we simulated data according to the seven-factor design from Simulation Study 1, with five replications per cell of the design. Each of the resulting 7,290 data matrices $\mathbf{X}$ was analyzed four times, each time with 25 random starts: (a) with K clusters and $Q - 1$ components per cluster, (b) with K clusters and $Q + 1$ components per cluster, (c) with $K - 1$ clusters and Q components per cluster, and (d) with $K + 1$ clusters and Q components per cluster—where K denotes the correct number of clusters and Q the correct number of components that is underlying

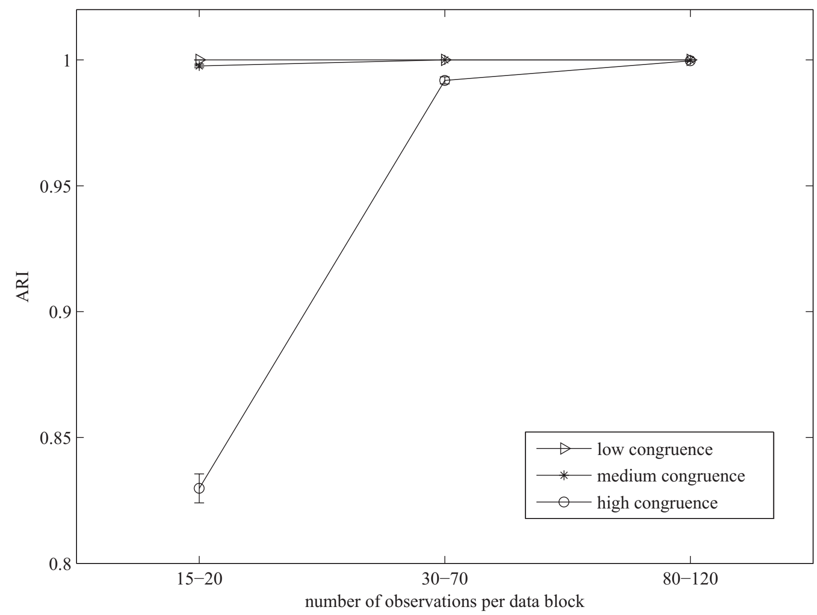

Figure 4. Mean adjusted Rand index (ARI) and associated 95% confidence intervals as a function of the number of observations per data block and the congruence of the cluster loading matrices (low, medium, or high).

the data. When using incorrect values of Q (Analyses a and b), the overall ARI equals .79 ($SD = .34$) and .96 ($SD = .16$) when $Q - 1$ and $Q + 1$ components are extracted for each cluster, respectively. Thus, extracting too many components (overextraction) appears to be less problematic for the recovery of the clustering than extracting too few components (underextraction). This makes sense, as for some data sets all Q components will be needed to properly distinguish between the clusters, whereas overextraction will most often yield all the information needed to disentangle the different clusters. Moreover, these results are in line with literature on the effects of overextraction and underextraction in component analysis (Fava & Velicer, 1992; Wood, Tataryn, & Gorsuch, 1996); that is, it is generally found that overextraction introduces less error to the estimated loading structure than underextraction.

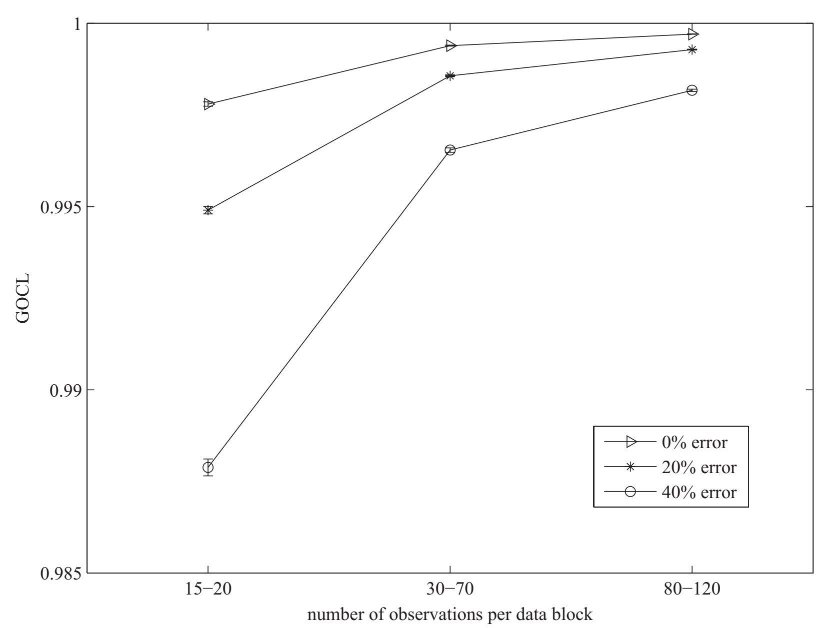

Figure 5. Mean goodness-of-cluster-loading-recovery (GOCL) value and associated 95% confidence intervals as a function of the number of observations per data block and the amount of error variance ($e \times 100\%$).

When using incorrect values of K, the most interesting question is how this affects the recovery of the clustering. When estimating too few clusters, we hypothesize that either two underlying clusters will be fused or the data blocks of one true cluster will be spread across two or more of the estimated clusters. Vice versa, when estimating too many clusters, we expect that either one true cluster will be split up or an additional cluster will be composed in the analysis out of an amalgam of data blocks that belong to different true clusters. To evaluate which of these phenomena occurs, we computed the ARI between the estimated clustering and the true clustering in which two clusters were merged (too low K used) or the true clustering splitting one of the true clusters into two clusters (too high K used); note that all possible splits and fusions were considered, retaining the split or fusion that yielded the highest ARI value. If these ARI values equal 1, this implies that the analysis deals with misspecification of K by cleanly fusing two true clusters or splitting one true cluster. Indeed, on average, ARI amounts to .99 (SD = .07) for the analyses with one cluster too few (note that the data sets with two true clusters were not taken into account) and .98 (SD = .08) for the analyses with one cluster too many.

Simulation Study 3

An additional small simulation study was performed to evaluate the goodness of recovery of clusterwise SCA–ECP analyses when the data contain variables that do not reflect the cluster structure of interest. These variables can mask the clustering and are thus named *masking variables* (see, e.g., Brusco & Cradit, 2001; Vichi & Kiers, 2001). The design of the simulation study consists of Factors 3, 4, 6, and 7 of Simulation Study 1, with the number of data blocks (Factor 1) fixed at 40, the number of observations per data block (Factor 2) between 80 and 120, and equal cluster sizes (Factor 5). Each cell of the design was replicated 10 times. To each of the thus obtained 810 data matrices, consisting of 12 variables, six masking variables were added. To test the influence of two kinds of masking variables, the six masking variables could be summarized either by the same two components across all data blocks (810 data sets in the no clustering condition) or by two components that differed across the data blocks and thus reflected another clustering of the data blocks than the clustering of interest (810 data sets in the conflicting-clustering condition). The loadings of these additional masking components are sampled uniformly between −1 and 1 (as in Level 1 of Factor 7 in Simulation Study 1). Each simulated data matrix is analyzed by clusterwise SCA–ECP with the correct number of clusters and components.

On average, the ARI between the true clustering and the estimated clustering amounts to .97 (SD = .13) and .71 (SD = .43) in the no-clustering and conflicting-clustering conditions, respectively. In the conflicting-clustering condition, the ARI largely depends on the level of congruence of the cluster loading matrices for the 12 original variables ($\hat{\omega}^2 = .50$): The mean ARI is .99, .86, and .29 for cluster loading matrices with low, medium, and high congruence, respectively.

Conclusion

On the basis of Simulation Study 1, we can conclude that for simulated data sets, the clusterwise SCA–ECP algorithm is not prone to end in a local minimum when using a multistart procedure with 25 random starts, given that the correct number of clusters and components are used. Furthermore, the algorithm appears to recover both the partition matrix and the relative sizes of the cluster loadings very well. The latter result puts the tendency of the algorithm for overfitting in perspective, as it suggests that the overfitting primarily boils down to a slight overestimation of the loadings, which is a well-known result in component analysis and does not affect the interpretation of the components (see, e.g., Velicer, Peacock, & Jackson, 1982). On the basis of Simulation Study 2, we can infer that the performance of clusterwise SCA–ECP is affected more by using too few clusters or components than by using too many. Typically, when using one cluster too many, one of the underlying clusters is split up in the estimated model. Finally, Simulation Study 3 has shown that masking variables have little influence on the goodness of recovery, except when the masking variables reflect a conflicting clustering and the clustering of interest is less distinct due to high congruence of the corresponding cluster-loading matrices.

The results of the simulation studies can be translated into some guidelines about the use of clusterwise SCA–ECP analysis in practice. First, Simulation Study 1 indicated that 25 random starts are sufficient under ideal conditions. Therefore, when analyzing empirical data, one may adopt the following strategy: When exploring different numbers of clusters and components, 25 starts can be used, but afterward, when one retains some solutions (i.e., some particular combinations of K and Q) for further investigation, it may be wise to repeat the estimation of these solutions with a larger number of random starts (say, 50 or 100) to reduce the probability of obtaining a local minimum. Also, Simulation Study 1 indicates that one should be cautious when the number of observations per data block is low, when the number of clusters and components may be large, when the data may contain a lot of error, and when the cluster loadings may be highly congruent. Among these data characteristics, the number of observations per data block is the only one that can be observed; therefore, we emphasize that it is important to have enough observations within each data block (preferably more observations than variables). Second, the results of Simulation Study 2 imply that when in doubt about the underlying number of clusters or components, it is advisable to consider high enough numbers. Third, on the basis of Simulation Study 3, we hypothesize that when the data contain masking variables, clusterwise SCA–ECP will still be able to recover the clustering, unless the structures of the clusters are highly congruent or the number of masking variables is too large.

Illustrative Application 1

In this section, we present an empirical example from eating disorder research. In this research domain, it has been observed that a substantial proportion of patients with eating disorders, like anorexia nervosa (AN) or bulimia nervosa (BN), engage in high levels of physical activity (Beumont, Arthur, Russell, & Touyz, 1994; Davis, 1997; Solenberger, 2001). Partly due to the absence of a clear operational definition of excessive exercise, few studies have provided estimates of the prevalence of this high-level exercising among eating disorder patients. However, there is some consensus that approximately 80% of the AN and 55% of the BN patients engage in excessive exercising (Davis, Kaptein, Kaplan, Olmsted, & Woodside, 1998; Epling, Pierce, & Stefan, 1983; Peñas-Lledó, Vaz Leal, & Waller, 2002). To explain this excessive physical activity at the within-patient level, two hypotheses have been put forward about the underlying psychological processes. The drive-for-thinness hypothesis states that eating disorder patients engage in physical activity because they are actively trying to lose weight by burning calories (Davis, 1997; Davis, Kennedy, Ravelski, & Dionne, 1994; Heatherton & Baumeister, 1991). The affect regulation hypothesis reads that physical activity is a way of coping with chronically negative affect (Davis, Katzman, & Kirsh, 1999; Holtkamp, Hebebrand, & Herpertz-Dahlmann, 2004; Thome & Espelage, 2004). In the subgroup of patients who do not display excessive levels of physical activity, physical activity at the within-patient level is not expected to be related to burning calories but to increased positive affect, just as in normal control subjects (Gauvin, Rejeski, & Norris, 1996; Kelsey et al., 2006).

To trace whether there are interindividual differences in how physical activity is related to drive for thinness and affect, Vansteelandt, Rijmen, Pieters, Probst, and Vanderlinden (2007) studied 32 female patients of the specialized inpatient eating disorder unit of the University Psychiatric Centre in Leuven, Belgium. Of these patients, 19 suffered from AN and 13 from BN. During 1 week, at nine randomly selected times a day the patients were signaled by an electronic device to fill out a questionnaire. The questionnaire consisted of 22 items that aimed at measuring the momentary drive for thinness, positive and negative emotional states, urge to be physically active, and physical activity. Inevitably, patients typically missed some of the 63 signals (7 days $\times$ 9 signals), resulting in a different number of assessments for different patients. On average 45.60 (SD = 10.78) observations were obtained per patient. This study results in a data matrix $\mathbf{X}$ of 1,459 observations by 22 variables, consisting of 32 submatrices $\mathbf{X}_i$, one for each patient, where each $\mathbf{X}_i$ was centered and autoscaled. Vansteelandt et al. also asked all patients to fill in the Eating Disorder Inventory, the Commitment to Exercise Scale, the Symptom Checklist, and the Beck Depression Inventory as between-patient measures.

To trace the most important interindividual differences in within-patient relations between drive for thinness, affect and physical activity, we analyzed the data set with clusterwise SCA–ECP, using 25 random starts and with K and Q varying from one to eight. The clusterwise SCA–ECP solutions with K equal to one are SCA–ECP solutions (see Model section). Separate PCAs were also performed. In Figure 6 the percentage of explained variance of the different clusterwise SCA–ECP solutions is plotted against the number of components for each value of K.

To select a model out of the obtained solutions, we used the model selection procedure that was described in the Data Analysis section. First, the number of clusters K is determined. Since going from one (i.e., SCA–ECP) to two clusters gives an increase in fit of about 2.5%, whereas adding more clusters hardly improves the percentage of explained variance, we decided to choose between the solutions with two clusters. Second, out of these solutions, a model with two components is chosen, since the scree line of the clusterwise SCA–ECP solutions with two clusters shows an elbow at that point. On top of that, the solution with two components and two clusters appeared to be the best in terms of interpretability of the cluster loading matrices and stability of the subject partition. For the eating disorder data, the best fitting partition allocates 25 subjects in Cluster 1 and seven subjects in Cluster 2. This partition

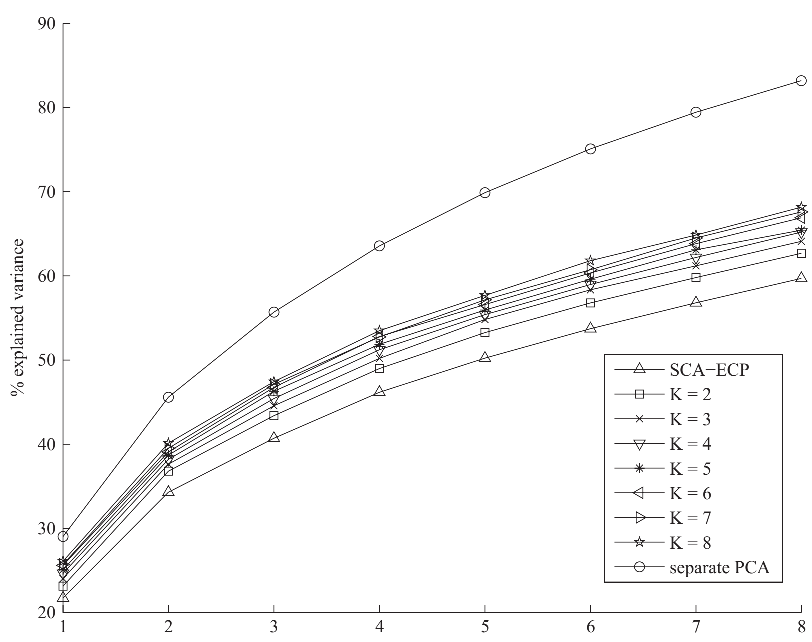

Figure 6. Percentage of explained variance for separate principal component analysis (PCA), simultaneous component analysis with equal average cross-products constraints (SCA–ECP), and clusterwise SCA–ECP solutions for the eating disorder data, with the number of components varying from one to eight and the number of clusters for clusterwise SCA–ECP varying from two to eight.

occurs once within the 25 runs. Out of the remaining runs, however, 22 yield a solution with a very similar partition in which only one or two subjects are assigned to another cluster. Reallocating these subjects does not alter the structure of the cluster loading matrices. Since the partition with seven subjects in Cluster 2, which is the global optimum to the best of our knowledge, is retrieved only once in the 25 randomly started runs of the algorithm, it may as well be missed in another 25 runs. So, in this case, performing a higher number of runs (e.g., 100) would be useful to decrease the probability of landing in a local minimum.

The normalized varimax rotated component loadings[2] for the two clusters of this solution are given in Table 5. In this table, it can be seen that in Cluster 1 all emotions load high on the first component, which is therefore labeled positive versus negative affect. Feeling fat and ugly also load high on this component and are related to negative affect. The second component of this cluster shows high loadings for the items measuring urge to be physically active, physical activity, and drive for thinness. This indicates that urge to be physically active, physical activity, and drive for thinness (especially drive for burning calories and losing weight) covary, which is in line with the drive-for-thinness hypothesis. In Cluster 2 a different structure shows up. The first component has highly negative loadings for the negative emotions and highly positive loadings for the emotions *pleased* and *happy*. Therefore this first component is labeled pleased versus negative affect. On the second component all positive emotional states are loading high, together with two of the physical activity items. The second component thus indicates that positive affect and physical activity are related for the subjects in Cluster 2.

In summary, the patients in the first cluster tend to engage in physical activity when they feel the momentary urge to be physically active and burn calories (drive for thinness). For the subjects in Cluster 2, however, physical activity is not driven by a drive for thinness or by affect regulation. Instead, physical activity is related to positive affect, so it seems that they tend to move more when they are feeling happier (or vice versa). As a result, it appears that clusterwise SCA–ECP made a distinction between the cluster of patients engaging in excessive activity to burn calories and the cluster of patients who do not engage in excessive physical activity and for whom activity is associated with positive affect as in normal controls. When we use the prevalence estimates of excessive exercising in eating disorders mentioned in the introduction of this section, we expect about 22 of the subjects (.80 × 25 AN patients + .55 × 7 BN patients) to belong to Cluster 1 and 10 to Cluster 2, which is in line with the results (25 belong to Cluster 1 and seven to Cluster 2).

These between-cluster differences in the mechanisms behind physical activity are further related to the severity of the eating disorder. Specifically, comparing the scores of the patients in Clusters 1 and 2 on the Beck Depression Inventory ($t = 4.36$, $p < .001$), the Commitment to Exercise Scale ($t = 3.48$, $p = .002$), and the Eating Disorder Inventory ($t = 2.66$, $p = .02$) reveals that the patients in Cluster 1 are more depressed, are more committed to physical exercise, and have higher body dissatisfaction. These results are in line with the original study of Vansteelandt et al. (2007), who also found that the within-patient relation between drive for thinness and urge to be physically active and physical activity—the relation found in Cluster 1—was stronger for patients scoring higher on depression in particular and pathology in general.

This application illustrates the value of clusterwise SCA–ECP as a generic modeling strategy, since it shows that with this strategy, a solution with a good balance between fit and model complexity can be obtained. Specifically, when two components are extracted, performing separate PCAs explains roughly 10% more variance than the standard (i.e., nonclusterwise) SCA–ECP approach. Yet, by classifying the 32 subjects in two clusters only, clusterwise SCA–ECP was already able to explain about 2.5% more variation than standard SCA–ECP. Since fitting more than two clusters leads to considerably smaller increases in fit, the additional clusters seem to chart more idiosyncratic structural differences. We can conclude that the selected clusterwise SCA–ECP solution is only slightly more complex than the corresponding SCA–ECP solution but yields much more information on the underlying mechanisms of physical activity for the eating disorder patients. More specifically, in line with earlier findings in eating disorder research, for the eating disorder patients who display high levels of (urge for) physical activity (Cluster 1), being active is associated with drive for thinness (Davis, 1997), whereas for the patients who do not engage in excessive physical activity (Cluster 2), activity is related to positive affect (Gauvin et al., 1996; Kelsey et al., 2006).

Illustrative Application 2

The second empirical example stems from social psychology and addresses how people react emotionally to unequal (and hence unfair) outcomes dependent on the personal need of another person. That is, people in need may arouse empathy, and this is likely to have an effect on the emotional reaction. This was examined with a public-good dilemma game in which the participant and an opposing player contribute to a public good. Note that a good becomes a public good when individuals can benefit from the good even if they did not contribute to its provision. Yet, when the combined investment of both players is sufficient, the public good can be provided and both obtain a reward. An important principle in public-good dilemmas is equality (van Dijk & Wilke, 1995, 2000): When both players contribute equally, they want to be rewarded equally. Violation of equality is perceived as unjust and elicits anger (Stouten, De Cremer, & van Dijk, 2005). In case of inequality, the subject's negative emotional reactions may be mollified if the advantaged person is in personal need, because feelings of sympathy and empathy can be aroused (Stouten, Ceulemans, Timmerman, & Van Hiel, 2011).

To study the effects of equality violation and the opponent being in need on a broad range of emotions, Stouten et al. (2011) performed an experimental study in which 282 participants played a public-good dilemma with an alleged opponent, in which it was said that sufficient contributions were made to provide the public good. Hence, both players earned a reward. In the experiment, two factors were fully crossed, yielding six conditions. First, the reward was equal for both players (equal conditions) or was higher for the opponent (unequal conditions). Second, the amount of empathy was manipulated at three levels. In the high-empathy

[2] Some eating disorder patients (especially in Cluster 2) displayed no variance on one or more of the variables. As the invariant scores cannot be rescaled to have a variance of 1, they were replaced by zeros. Therefore, in this case, even though the components are orthogonal, the loadings should be read as weights rather than correlations.

Table 5
Normalized Varimax Rotated Loadings for the Two Clusters of the Eating Disorder Data

Variable	Cluster 1: Positive affect vs. negative affect	Cluster 1: Drive-for-thinness hypothesis	Cluster 2: Pleased vs. negative affect	Cluster 2: Positive affect and Physical activity
		Positive affect		
Pleased	**.71**	.09	**.47**	**.45**
Happy	**.72**	.07	**.34**	**.53**
Appreciated	**.48**	.08	−.08	**.73**
Love	**.41**	.12	.09	**.69**
		Negative affect		
Sad	**−.76**	−.05	**−.75**	.01
Angry	**−.69**	.00	**−.32**	.03
Lonely	**−.65**	−.01	**−.68**	−.04
Anxious	**−.59**	.06	**−.30**	−.06
Tense	**−.66**	.10	**−.51**	−.22
Guilty	**−.55**	.04	**−.31**	.02
Irritated	**−.52**	.06	**−.36**	−.08
Ashamed	**−.53**	.01	−.09	.05
		Drive for thinness		
Feel fat	**−.43**	.28	−.15	−.19
Feel ugly	**−.46**	.18	−.23	.01
Want to burn calories	−.19	**.61**	.00	−.04
Want to lose weight	−.30	**.34**	−.08	.01
		Urge to be physically active		
Want to be active	−.07	**.78**	.11	−.12
Want to move	−.07	**.77**	.12	−.11
Want to sport	−.07	**.73**	−.02	−.09
		Physical activity		
Am active	.15	**.61**	−.03	**.46**
Am moving	.11	**.61**	−.06	**.51**
Am sporting	.16	**.48**	−.02	.17

Note. Loadings greater than ±.30 are highlighted in bold.

conditions, participants received a message about a negative personal event that happened to the opponent and were asked to imagine how the opponent felt. In the low-empathy conditions, participants also received the above message but were asked to take an objective perspective. In the control conditions, participants did not receive a message. After the game, participants were asked how they felt using 21 emotion terms, to be rated on a 7-point Likert scale. This study resulted in a data matrix of 282 participants by 21 emotions, where each of the participants was nested within a particular equality or empathy condition.

To investigate how the structure of the experienced emotions within a condition possibly differs across the six conditions, we applied clusterwise SCA–ECP to the data. We used 25 random starts and varied K from one to six and Q from one to six. Clusterwise SCA–ECP solutions with K equal to one are SCA–ECP solutions, and clusterwise SCA–ECP with K equal to six boils down to the separate PCA approach, since there are six conditions (see Model section). In Figure 7, the percentage of explained variance of the different clusterwise SCA–ECP solutions is plotted against the number of components for each value of K. Upon inspection of Figure 7, we decided to select a solution in which the conditions are grouped into two clusters, because going from one (SCA–ECP) to two clusters gives a large increase in fit, whereas adding more clusters gives substantially smaller improvements in fit. Subsequently, we selected the model with two components, since the scree line of the clusterwise SCA–ECP solutions with two clusters showed a clear elbow at two components.

In the selected solution, Cluster 1 contains all equal conditions (equal/high empathy, equal/low empathy, equal/control) and Cluster 2 all unequal conditions (unequal/high empathy, unequal/low empathy, unequal/control). Thus, the equality factor seems to have the largest effect on the underlying structure of the emotion patterns in the different conditions.

The normalized varimax rotated component loadings for the two clusters of the selected solution are given in Table 6. In this table, it can be seen that the first component of the cluster with the equal conditions is a positive affect versus negative affect component. The second component is characterized by high loadings for empathic emotions such as *compassionate* and *tenderhearted* and thus can be labeled empathy. The components of the unequal conditions are rather similar to the ones of the equal conditions and can be labeled positive versus negative affect and empathy as well. However, a very different loading pattern across the two clusters is found for three emotions: *fearful*, *surprised*, and *sympathetic*.

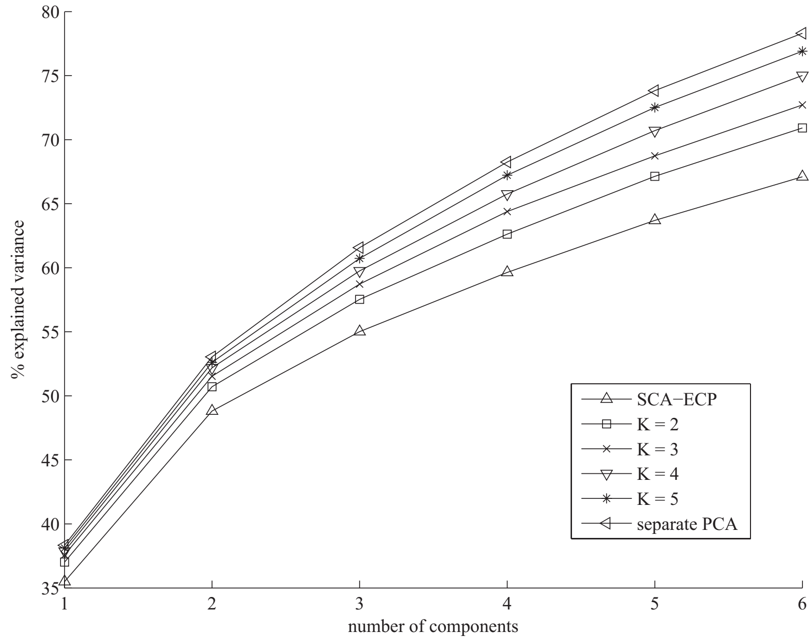

Figure 7. Percentage of explained variance for separate principal component analysis (PCA), simultaneous component analysis with equal average cross-products constraints (SCA–ECP), and clusterwise SCA–ECP solutions for the dilemma game data, with the number of components varying from one to six and the number of clusters for clusterwise SCA–ECP varying from two to five.

Fearful loads highly on the positive versus negative affect component in the equal conditions cluster, whereas in the cluster with unequal conditions it is related to the empathy emotions of the second component. A tentative interpretation of this reads that within the unequal conditions, it may be that people who react empathically (e.g., based on the message received) to the unequal treatment also feel fearful because they do not know what is going to happen next. Additionally, participants may feel powerless, which is also argued to be one of the core components of fear (Ellsworth & Smith, 1988). Hence, due to the uncertainty of the situation and the fact that participants have to accept the given situation, they also feel fearful.

Surprised loads high on the empathy component in the equal conditions and on the positive versus negative affect component in the unequal conditions. Thus, in the equal conditions the feelings of surprise go together with an empathic reaction, whereas in the unequal conditions surprise is associated with anger about the unequal treatment of the players of the game. This negative surprise is elicited as part of the unexpected result of receiving unequal outcomes (e.g., feeling stunned, seeking understanding).

Concerning *sympathetic*, it can be seen that in the unequal conditions this emotion loads higher on the empathy component than in the equal conditions. Although participants received an unequal (and unfair) outcome, the fact that the other person was in need elicits understanding and therefore the feeling of sympathy. This implies that in the unequal conditions, sympathy is associated with an empathetic reaction toward the opponent, whereas this is less outspoken in the equal conditions.

This empirical example with subjects nested within groups shows that clusterwise SCA–ECP can also be a useful tool to explore structural differences between groups of subjects. Moreover, this application demonstrates that clusterwise SCA–ECP is able to detect interpretable differences in component structure, even when these differences are small.

Discussion

In scientific research, many questions concern differences and similarities in the underlying structure of multivariate multiblock data. To answer such research questions, two commonly used strategies are separate principal component analysis and simultaneous component analysis, with the first one often being too flexible to provide insight and the second one being too parsimonious to detect differences. Therefore, the clusterwise SCA–ECP model was introduced as a generic modeling strategy for detecting structural differences and similarities in multiblock multivariate data, where each data block can consist of multiple measurements for a certain subject or of multiple subjects that belong to a particular group. Specifically, the most important structural differences between the data blocks are captured by clustering the data blocks: Data blocks with similar component loadings are grouped in the same cluster, whereas data blocks from different clusters have different component loadings. Thus, the structural differences between the data blocks can be derived from the comparison of the cluster loading matrices. Clusterwise SCA–ECP comprises the separate PCA (number of clusters equal to number of data blocks) and SCA–ECP (number of clusters equal to one) approaches as special cases, as well as all the intermediate solutions (number of clusters between one and number of data blocks). Therefore, it allows one to look for a solution with a good balance between fit and model complexity.

Table 6
Normalized Varimax Rotated Loadings for the Two Clusters of the Dilemma Game Data

Emotion	Equal conditions: Positive affect vs. negative affect	Equal conditions: Empathy	Unequal conditions: Positive affect vs. negative affect	Unequal conditions: Empathy
Angry	**−.84**	−.06	**−.76**	.00
Hurt	**−.77**	.01	**−.47**	.20
Irritated	**−.84**	.14	**−.79**	−.02
Annoyed	**−.85**	.10	**−.87**	−.10
Frustrated	**−.85**	.15	**−.83**	−.10
Disappointed	**−.84**	.09	**−.75**	−.00
Indignant	**−.81**	.15	**−.76**	−.20
Enraged	**−.80**	.10	**−.74**	−.05
Hostile	**−.75**	−.04	**−.61**	−.01
Fearful	**−.77**	.16	−.17	**.49**
Surprised	−.29	**.46**	**−.50**	.01
Relieved	.28	**.54**	**.43**	**.51**
Elated	**.57**	**.39**	.27	**.62**
Warm	**.44**	**.47**	.30	**.51**
Happy	**.72**	.28	**.64**	.35
Satisfied	**.77**	.23	**.63**	.31
Sympathetic	.16	.25	.03	**.62**
Compassionate	−.34	**.57**	.02	**.61**
Tenderhearted	−.21	**.72**	−.07	**.74**
Concerned	**−.37**	**.52**	−.14	**.55**
Tender	.09	**.76**	.22	**.42**

Note. Loadings greater than ± .35 are highlighted in bold.

We see various directions for future research, at the level of the data, the level of the model, and the level of the data analysis.

Level of the Data

The data blocks were autoscaled, implying that the information on the between-block differences in variability is lost. For certain research questions, these differences can be very interesting; for example, to study whether emotions vary more over time for some people than for others and whether these differences in variability are related to personality traits (Oravecz, Tuerlinckx, & Vandekerckhove, 2009). Therefore, it may be worthwhile to consider standardizing the variables over the data blocks. This implies severe adjustments to the modeling strategy, however.

Level of the Model

By using SCA–ECP within each cluster, all between-block differences in correlations between component scores must go into the clustering. Thus, data blocks with the same component loadings but different correlations between the components would be assigned to separate clusters. For example, when personality traits are measured for subjects from a number of countries, it is possible that although in some countries the same personality traits are found, they are differentially correlated. In that case, clusterwise SCA–ECP would assign these countries to separate clusters, whereas putting them in the same cluster would be more parsimonious. A solution would be to use a variant of SCA within the clusters that allows for differences in correlations between the components.

Level of the Data Analysis

In this article, the number of components was fixed across the clusters. This is often not very realistic. For example, in studies on emotional granularity, subjects with a high granularity are expected to have more components in their affect space than subjects with a low emotional granularity. Thus, in future research it would be useful to allow the number of components to vary between clusters.

Conclusion

We presented clusterwise SCA–ECP, a versatile and generic modeling strategy for detecting and describing structural differences and similarities in multivariate multiblock data. User-friendly clusterwise SCA–ECP software is available online (http://ppw.kuleuven.be/okp/software/MBCA/). Based on simulation studies, guidelines were formulated for the use of clusterwise SCA–ECP analysis in practice.

References

Aloise, D., Hansen, P., & Liberti, L. (2010). An improved column generation algorithm for minimum sum-of-squares clustering. *Mathematical Programming.* Advance online publication. doi:10.1007/s10107-010-0349-7

Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & Emotion, 12,* 579–599. doi:10.1080/026999398379574

Beumont, P. J. V., Arthur, B., Russell, J. D., & Touyz, S. W. (1994). Excessive physical activity in dieting disorder patients: Proposals for a supervised exercise program. *International Journal of Eating Disorders, 15,* 21–36. doi:10.1002/1098-108X(199401)15:1<21::AID-EAT2260150104>3.0.CO;2-K

Bock, H. H. (1987). On the interface between cluster analysis, principal component analysis, and multidimensional scaling. In H. Bozdogan & A. K. Gupta (Eds.), *Multivariate statistical modeling and data analysis* (pp. 17–34). Dordrecht, the Netherlands: Reidel.

Bouveyron, C., Girard, S., & Schmid, C. (2007). High-dimensional data clustering. *Computational Statistics & Data Analysis, 52,* 502–519. doi:10.1016/j.csda.2007.02.009

Brusco, M. J. (2006). A repetitive branch-and-bound procedure for minimum within-cluster sums of squares partitioning. *Psychometrika, 71,* 347–363. doi:10.1007/s11336-004-1218-1

Brusco, M. J., & Cradit, J. D. (2001). A variable selection heuristic for *K*-means clustering. *Psychometrika, 66,* 249–270. doi:10.1007/BF02294838

Brusco, M. J., & Cradit, J. D. (2005). ConPar: A method for identifying groups of concordant subject proximity matrices for subsequent multidimensional scaling analyses. *Journal of Mathematical Psychology, 49,* 142–154. doi:10.1016/j.jmp.2004.11.004

Brusco, M. J., Cradit, J. D., Steinley, D., & Fox, G. L. (2008). Cautionary remarks on the use of clusterwise regression. *Multivariate Behavioral Research, 43,* 29–49. doi:10.1080/00273170701836653

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245–276. doi:10.1207/s15327906mbr0102_10

Davis, C. (1997). Eating disorders and hyperactivity: A psychobiological perspective. *Canadian Journal of Psychiatry, 42,* 168–175.

Davis, C., Kaptein, S., Kaplan, A. S., Olmsted, M. P., & Woodside, D. B. (1998). Obsessionality in anorexia nervosa: The moderating influence of exercise. *Psychosomatic Medicine, 60,* 192–197.

Davis, C., Katzman, D. K., & Kirsh, C. (1999). Compulsive physical activity in adolescents with anorexia nervosa: A psychobehavioral spiral of pathology. *Journal of Nervous and Mental Disease, 187,* 336–342. doi:10.1097/00005053-199906000-00002

Davis, C., Kennedy, S. H., Ravelski, E., & Dionne, M. (1994). The role of physical activity in the development and maintenance of eating disorders. *Psychological Medicine, 24,* 957–967. doi:10.1017/S0033291700029044

De Roover, K., Ceulemans, E., & Timmerman, M. E. (2011). How to perform multiblock component analysis in practice. *Behavior Research Methods.* Advance online publication. doi:10.3758/s13428-011-0129-1

DeSarbo, W. S., Oliver, R. L., & Rangaswamy, A. (1989). A simulated annealing methodology for clusterwise linear regression. *Psychometrika, 54,* 707–736. doi:10.1007/BF02296405

de Soete, G., & Carroll, J. D. (1994). *K*-means clustering in a low-dimensional Euclidean space. In E. Diday, Y. Léchevallier, M. Schader, P. Bertrand, & B. Burtschy (Eds.), *New approaches in classification and data analysis* (pp. 212–219). Berlin, Germany: Springer.

Dolan, C. V., Oort, F. J., Stoel, R. D., & Wicherts, J. M. (2009). Testing measurement invariance in the target rotated multigroup exploratory factor model. *Structural Equation Modeling, 16,* 295–314. doi:10.1080/10705510902751416

Dolan, C. V., & van der Maas, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika, 63,* 227–253. doi:10.1007/BF02294853

Eid, M., & Diener, E. (2001). Norms for experiencing emotions in different cultures: Inter- and intranational differences. *Journal of Personality and Social Psychology, 81,* 869–885. doi:10.1037/0022-3514.81.5.869

Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 45–60). Chichester, England: Wiley.

Ellsworth, P. C., & Smith, C. A. (1988). From appraisal to emotion: Differences among unpleasant feelings. *Motivation and Emotion, 12,* 271–302. doi:10.1007/BF00993115

Epling, W. F., Pierce, W. D., & Stefan, L. (1983). A theory of activity-based anorexia. *International Journal of Eating Disorders, 3,* 27–46. doi:10.1002/1098-108X(198323)3:1<27::AID-EAT2260030104>3.0.CO;2-T

Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research, 27,* 387–415. doi:10.1207/s15327906mbr2703_5

Fontaine, J. R. J., Poortinga, Y. H., Setiadi, B., & Markam, S. S. (2002). Cognitive structure of emotion terms in Indonesia and the Netherlands. *Cognition & Emotion, 16,* 61–86. doi:10.1080/02699933014000130

Fontaine, J. R. J., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science, 18,* 1050–1057. doi:10.1111/j.1467-9280.2007.02024.x

Gauvin, L., Rejeski, W. J., & Norris, J. L. (1996). A naturalistic study of the impact of acute physical activity on feeling states and affect in women. *Health Psychology, 15,* 391–397. doi:10.1037/0278-6133.15.5.391

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica, 11,* 1–12. doi:10.2307/1905714

Hands, S., & Everitt, B. (1987). A Monte Carlo study of the recovery of cluster structure in binary data by hierarchical clustering techniques. *Multivariate Behavioral Research, 22,* 235–243. doi:10.1207/s15327906mbr2202_6

Hays, W. L. (1963). *Statistics for psychologists.* New York, NY: Holt, Rinehart & Winston.

Heatherton, T. F., & Baumeister, R. F. (1991). Binge eating as escape from self-awareness. *Psychological Bulletin, 110,* 86–108. doi:10.1037/0033-2909.110.1.86

Hessen, D. J., Dolan, C. V., & Wicherts, J. M. (2006). Multigroup common factor model with minimal uniqueness constraints and the power to detect uniform bias. *Applied Psychological Measurement, 30,* 233–246. doi:10.1177/0146621605279760

Holtkamp, K., Hebebrand, J., & Herpertz-Dahlmann, B. (2004). The contribution of anxiety and food restriction on physical activity levels in acute anorexia nervosa. *International Journal of Eating Disorders, 36,* 163–171. doi:10.1002/eat.20035

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2,* 193–218. doi:10.1007/BF01908075

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). STEMM: A general finite mixture structural equation model. *Journal of Classification, 14,* 23–50. doi:10.1007/s003579900002

Jolliffe, I. T. (1986). *Principal component analysis.* New York, NY: Springer.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36,* 409–426. doi:10.1007/BF02291366

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23,* 187–200. doi:10.1007/BF02289233

Kelsey, K. S., DeVellis, B. M., Begum, M., Belton, L., Hooten, E. G., & Campbell, M. K. (2006). Positive affect, exercise and self-reported health in blue-collar women. *American Journal of Health Behavior, 30,* 199–207.

Kiers, H. A. L. (1990). *SCA: A program for simultaneous components analysis of variables measured in two or more populations.* Groningen, the Netherlands: IEC ProGAMMA.

Kiers, H. A. L., & ten Berge, J. M. F. (1994). Hierarchical relations between methods for simultaneous component analysis and a technique for rotation to a simple simultaneous structure. *British Journal of Mathematical and Statistical Psychology, 47,* 109–126. doi:10.1111/j.2044-8317.1994.tb01027.x

Kline, R. B. (2004). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.

Kroonenberg, P. M. (2008). *Applied multiway data analysis.* Hoboken, NJ: Wiley. doi:10.1002/9780470238004

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling.* Newbury Park, CA: Sage.

Lawley, D. N., & Maxwell, A. E. (1962). Factor analysis as a statistical method. *The Statistician, 12,* 209–229. doi:10.2307/2986915

MacKinnon, N. J., & Keating, L. J. (1989). The structure of emotions: Canada–United States comparisons. *Social Psychology Quarterly, 52,* 70–83. doi:10.2307/2786905

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). Berkeley: University of California Press.

McLachlan, G., & Peel, D. (2000). *Finite mixture models.* New York, NY: Wiley.

Meredith, W., & Millsap, R. E. (1985). On component analyses. *Psychometrika, 50,* 495–507. doi:10.1007/BF02296266

Milligan, G. W., Soon, S. C., & Sokol, L. M. (1983). The effect of cluster size, dimensionality, and the number of clusters on recovery of true cluster structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 5,* 40–47. doi:10.1109/TPAMI.1983.4767342

Millsap, R. E., & Meredith, W. (1988). Component analysis in cross-sectional and longitudinal data. *Psychometrika, 53,* 123–134. doi:10.1007/BF02294198

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology, 25,* 241–286. doi:10.1006/ceps.2000.1040

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2009). A hierarchical Ornstein–Uhlenbeck model for continuous repeated measurement data. *Psychometrika, 74,* 395–418. doi:10.1007/s11336-008-9106-8

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 2,* 559–572.

Peñas-Lledó, E., Vaz Leal, F. J., & Waller, G. (2002). Excessive exercise in anorexia nervosa and bulimia nervosa: Relation to eating characteristics and general psychopathology. *International Journal of Eating Disorders, 31,* 370–375. doi:10.1002/eat.10042

Rocci, R., & Vichi, M. (2005). Three-mode component analysis with crisp or fuzzy partition of units. *Psychometrika, 70,* 715–736. doi:10.1007/s11336-001-0926-z

Rodríguez, C., & Church, A. T. (2003). The structure and personality correlates of affect in Mexico: Evidence of cross-cultural comparability using the Spanish language. *Journal of Cross-Cultural Psychology, 34,* 211–230. doi:10.1177/0022022102250247

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology, 76,* 805–819. doi:10.1037/0022-3514.76.5.805

Schouteden, M., Van Deun, K., Van Mechelen, I., & Pattyn, S. (2010). *SCA and rotation to distinguish common and specific information in coupled data.* Manuscript submitted for publication.

Solenberger, S. E. (2001). Exercise and eating disorders: A 3-year inpatient hospital record analysis. *Eating Behaviors, 2,* 151–168. doi:10.1016/S1471-0153(01)00026-5

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27,* 229–239. doi:10.1111/j.2044-8317.1974.tb00543.x

Späth, H. (1979). Algorithm 39: Clusterwise linear regression. *Computing, 22,* 367–373. doi:10.1007/BF02265317

Späth, H. (1982). Algorithm 48: A fast algorithm for clusterwise linear regression. *Computing, 29,* 175–181. doi:10.1007/BF02249940

Steinley, D. (2003). Local optima in *K*-means clustering: What you don't know may hurt you. *Psychological Methods, 8,* 294–304. doi:10.1037/1082-989X.8.3.294

Stouten, J., Ceulemans, E., Timmerman, M. E., & Van Hiel, A. (2011). Tolerance of justice violations: The effects of need on emotional reactions after violating equality in social dilemmas. *Journal of Applied Social Psychology, 41,* 357–380. doi:10.1111/j.1559-1816.2010.00717.x

Stouten, J., De Cremer, D., & van Dijk, E. (2005). All is well that ends well, at least for proselfs: Emotional reactions to equality violation as a function of social value orientation. *European Journal of Social Psychology, 35,* 767–783. doi:10.1002/ejsp.276

ten Berge, J. M. F. (1993). *Least squares optimization in multivariate analysis.* Leiden, the Netherlands: DSWO Press.

Thome, J., & Espelage, D. L. (2004). Relations among exercise, coping, disordered eating, and psychological health among college students. *Eating Behaviors, 5,* 337–351. doi:10.1016/j.eatbeh.2004.04.002

Timmerman, M. E., Ceulemans, E., Kiers, H. A. L., & Vichi, M. (2010). Factorial and reduced *K*-means reconsidered. *Computational Statistics & Data Analysis, 54,* 1858–1871. doi:10.1016/j.csda.2010.02.009

Timmerman, M. E., & Kiers, H. A. L. (2003). Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika, 68,* 105–121. doi:10.1007/BF02296656

Timmerman, M. E., Kiers, H. A. L., Smilde, A. K., Ceulemans, E., & Stouten, J. (2009). Bootstrap confidence intervals in multi-level simultaneous component analysis. *British Journal of Mathematical and Statistical Psychology, 62,* 299–318. doi:10.1348/000711007X265894

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* (Personnel Research Section Report No. 984). Washington, DC: U.S. Department of the Army.

Tucker, L. R., & Messick, S. (1963). An individual differences model for multidimensional scaling. *Psychometrika, 28,* 333–367. doi:10.1007/BF02289557

Tugade, M. M., Fredrickson, B. L., & Barrett, L. F. (2004). Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on coping and health. *Journal of Personality, 72,* 1161–1190. doi:10.1111/j.1467-6494.2004.00294.x

Van Deun, K., Smilde, A. K., van der Werf, M. J., Kiers, H. A. L., & Van Mechelen, I. (2009). A structured overview of simultaneous component based data integration. *BMC Bioinformatics, 10,* 246. doi:10.1186/1471-2105-10-246

van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research.* Thousand Oaks, CA: Sage.

van Dijk, E., & Wilke, H. (1995). Coordination rules in asymmetric social dilemmas: A comparison between public good dilemmas and resource dilemmas. *Journal of Experimental Social Psychology, 31,* 1–27. doi:10.1006/jesp.1995.1001

van Dijk, E., & Wilke, H. (2000). Decision-induced focusing in social dilemmas: Give-some, keep-some, take-some, and leave-some dilemmas. *Journal of Personality and Social Psychology, 78,* 92–104. doi:10.1037/0022-3514.78.1.92

Vansteelandt, K., Rijmen, F., Pieters, G., Probst, M., & Vanderlinden, J. (2007). Drive for thinness, affect regulation and physical activity in eating disorders: A daily life study. *Behaviour Research and Therapy, 45,* 1717–1734. doi:10.1016/j.brat.2006.12.005

Varriale, R., & Vermunt, J. K. (2010). *Multilevel mixture factor models.* Manuscript submitted for publication.

Velicer, W. F., Peacock, A. C., & Jackson, D. N. (1982). A comparison of component and factor patterns: A Monte Carlo approach. *Multivariate Behavioral Research, 17,* 371–388. doi:10.1207/s15327906mbr1703_5

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology, 33,* 213–239. doi:10.1111/j.0081-1750.2003.t01–1-00131.x

Vermunt, J. K. (2008a). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research, 17,* 33–51. doi:10.1177/0962280207081238

Vermunt, J. K. (2008b). Multilevel latent variable modeling: An application in education testing. *Austrian Journal of Statistics, 37,* 285–299.

Vichi, M., & Kiers, H. A. L. (2001). Factorial *K*-means analysis for two-way data. *Computational Statistics & Data Analysis, 37,* 49–64. doi:10.1016/S0167-9473(00)00064-5

Vichi, M., Rocci, R., & Kiers, H. A. L. (2007). Simultaneous component and clustering models for three-way data: Within and between approaches. *Journal of Classification, 24,* 71–98. doi:10.1007/s00357-007-0006-x

Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1,* 354–365. doi:10.1037/1082-989X.1.4.354

Yung, Y.-F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika, 62,* 297–330. doi:10.1007/BF02294554

Received December 16, 2009
Revision received May 1, 2011
Accepted May 17, 2011 ■